A Putative Photolyase from Haloferax volcanii

Wu, Winnie

Award date:
2009

Awarding institution:
University of Bath

A Putative Photolyase from *Hfx. volcanii*

Winnie Wu

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

2009

Acknowledgements

Summing up my PhD in one page simply doesn't do it justice. This is definitely one of the best life lessons one can receive, and I honestly never thought that I would get this far, both academically and geographically. This page marks the end of my thesis, which means that I need to give thanks to the people that have made this experience all worthwhile.

First and foremost, I would like to express my utmost gratitude to Dr. David Hough and Dr. Momna Hejmadi, two of the best PhD supervisors one can have. Their patience and support throughout these years have given me the will and courage to keep up with my work. I am particularly thankful to both for their valuable guidance and supervision throughout these years.

I would like to thank Dr. Dina Musaid AL-Mailem (University of Kuwait) for providing me with the essential knowledge in growing these pink bugs. I would also like to thank the wonderful staff at the teaching lab for lending me a key and trusting me enough not to lose it for the whole 2 years.

To the Extremophiles crew, past and present, I would like to thank each of you, as you all have provided me with help in some way or another. The list includes: Cal, Dina, Harry, Sarah, Mareike, Charlie, Natasha, Philippe, Vicki, Caroline, Jon, Nia, Dalal, Alex, Daniel, Chris, Carolyn, Karl, Sylvain, Tracey, Jodie and Albert (Phew!). Many of you have provided me with numerous laughs over the years, so thank you for adding to the entertainment value of my PhD. A special thanks goes to Charlie whose final year report proved extremely useful and whose pep talks have helped me get through the hard times. And Karl, the nicest and most helpful person I know, has made a difference to my PhD experience and I am grateful for that (and thanks for reading my thesis, too!).

I would also like to extend my thanks to my assessor Professor Michael Danson for his help and contribution to my project over the years. I'd also like to thank Dr. Susan Crennell and all members of the Centre for Extremophiles Research for their advice and help during my PhD. And of course, I must thank the University of Bath for giving me the opportunity to do my PhD in this wonderful city.

I am very grateful to mum, dad and James for their constant support, love and encouragement. And to all my lovely friends back in Vancouver and Taiwan, I would like to thank you for all the phone calls, e-mails and presents over the years!

Also, I need to express my gratitude to my fellow RTs for their understanding when I needed to swap my duties. And finally, I would like to thank Charles for being there and for listening to my complaints, all the time. Hopefully I have not forgotten to thank anyone in this acknowledgement!

Table of Contents

List of Figures

Abbreviations

8-HDF	8-hydroxy-5-deazaflavin
bp	base pair
BSA	bovine serum albumin
CPD	cyclobutane pyrimidine dimer
Cry	cryptochrome
Cry-DASH	cryptochrome (Drosophila, Arabidopsis, Synechocystis and Human)
DHlipDH	dihydrolipoamide dehydrogenase
DMSO	dimethyl sulphoxide
DNA	deoxynucleic acid
dNTP	deoxynucleoside triphosphate
DSMZ	Deutsche Sammulung von Mikroorganismen und Zellkulturen GmbH Braunschweig, Germany (German Collection of Microorganisms and Cell Culture)
DTT	dithiothreitol
E3	dihydrolipoamide dehydrogenase
EDTA	ethylenediamine tetracetic acid, (disodium)
FAD	flavin adenine dimucleotide
Fdx	ferredoxin
GSH	reduced glutathione
GSSG	oxidized glutathione
Hfx. volcanii	*Haloferax volcanii*
H98	*Hfx. volcanii* H98 expression strain (ΔpyrE2 ΔhdrB)
Hr	Hour
Hv-YPC	complete medium for H98 (Yeast, Peptone, and Casamino acids)
IOD	integral optical density
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilo base
LB	Luria-Bertani
LGT	lateral gene transfer
Min	minute
MTHF	5,10-methenyltetrahydrofolate

M_r	relative molecular mass
NER	nucleotide excision repair
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
Pyr<>Pyr	cyclobutane pyrimidine dimer
Pyr [6-4] pyr	pyrimidine-pyrimidone (6-4) photoproduct
RNA	ribonucleic acid
rpm	revolutions per minute
s	seconds
SDS	sodium dodecyl sulphate
TAE	tris-acetate EDTA
UV	ultraviolet
X-Gal	5-bromo-4-chloro-3-indoyl β-D-galactopyranoside

Abstract

Extreme halophilic archaea, like *Haloferax volcanii*, which thrive in hypersaline environments, are subject to intense UV irradiation in their natural habitats; however, the mechanisms underlying their superb tolerance to UV-induced cytotoxic DNA lesions, mainly cyclobutane pyrimidine dimers (Pyr<>Pyr) and pyrimidine – pyrimidone (6-4) photoproducts (Pyr [6-4] pyr), remain unclear. Among the various DNA repair pathways used by halophilic archaea, photoreactivation is the simplest and most efficient reversal mechanism. In this light-dependent process, the monomeric enzyme, photolyase, repairs the DNA lesion and causes the splitting of the cyclobutane ring via the absorption of visible light. This mechanism has been observed in several species of halophilic archaea, where light repair results in higher cell survival than dark repair.

Annotation of the complete genome sequence of the UV-tolerant haloarcheon, *Halobacterium* sp. NRC-1 has identified 2 photolyase-like genes, *phr 1* and *phr 2* where knock-out studies have confirmed the function of the encoded Phr2 protein as a CPD photolyase. However, this protein has not been expressed and functionally characterized from any halophilic archaeon known to date. With this in mind, a project was designed to investigate the expression and function of the *phr* encoded proteins from *Haloferax volcanii* with the aim to elucidate their function in DNA repair. *Haloferax volcanii* is a halophilic archaeon that grows optimally at 42°C under pH7, and requires an optimum salt concentration between 1.5M -2.5M.

Using the known *phr* gene sequences from *Halobacterium* sp. NRC-1, two putative gene homologues, *phr1* and *phr2*, were identified from the unannotated *Hfx. volcanii* genome. Homology modelling to known photolyase structures of *Escherichia coli* and *Anacystis nidulans* was conducted where the hypothetical structure of the protein closely resembled a deavaflavin type of photolyase containing the 8-HDF chromophore. The putative genes were amplified and transformed into the native strain to observe an over-expression of the Phr proteins. Due to the low copy number of the vector used, protein expression was too low to be detected; this led to a modification scheme for the vector to include a polyhistidine tag upstream of the gene for Nickel affinity purification. This strategy resulted in the purification of a native protein, pitA, a reflection of the low expression of Phr2 in the native strain. Heterologous expression in *E. coli* hosts resulted in the production of inclusion bodies that required renaturation to re-gain the protein native conformation. The Phr2 protein was purified, solubilised using 8M urea and refolded using various tested refolding

buffers. Circular dichroism spectroscopy revealed that the refolded protein has secondary structures with a high ratio of α-helixes.

A novel *in-vivo* assay developed for detecting the UV-sensitivity of the native overexpression strain strongly confirmed a photoreactivating function of the *phr2*-encoded protein but not for the *phr1*-encoded gene product. A cell viability assay indicated that white light illumination resulted in higher cell numbers for the Phr2 strain post UV irradiation. Lastly, an immunoassay was used for the detection of photoreactivating activity of the re-folded protein and the native over-expression strain. Results indicated that while the re-folded protein did not exhibit repair, the native over-expression strain once again showed a white light enhancement of repair of dimers.

These results and findings are discussed in context with the current literature on photoreactivation and DNA repair in other halophilic archaea.

Introduction

1.1 Phylogeny and characteristics of archaea

In 1962, Stanier and van Niel put forward the concept that all living organisms can be classified into either eukaryotes or prokaryotes based on differences in morphology, physiology and pathogenicity (Stanier and Van Niel 1962); however, many argued this classical view of taxonomy was inadequate in that the two groups did not represent true phylogentic distinctions and that it failed to establish meaningful evolutionary relationships used to group species into higher taxonomic orders. Owing to its remarkable degree of sequence conservation across all organisms, the gene encoding the small subunit ribosomal RNA (SSU rRNA) was used for the new classification system as proposed by Woese (Woese and Fox 1977). Using this emerging technique of nucleic acid sequencing, Woese found that a group of methane-generating anaerobic bacteria could not be classified under either of the two categories; hence, a new group of organisms, the Archaebacteria, was recommended to be given equal footing to both eukaryotes and bacteria (formerly Eubacteria) (Woese et al. 1990). They suggested that prokaryotes were much more diverse than originally thought, and that the phylogenetic structure needed to be re-examined.

It is now widely recognized that three evolutionarily distinct domains exist: the Bacteria, the Archaea, and the Eukarya (Figure 1.1). The discovery of eukaryotic traits in Archaea coupled with phylogenetic analyses using paralogue proteins led to the proposal that the root for the tree of life lies in the Bacterial domain (Zillig et al. 1985; Gogarten et al. 1989; Iwabe et al. 1989; Brown and Doolittle 1995; Lawson et al. 1996; Brown et al. 1997; Gribaldo and Cammarano 1998; Labedan et al. 1999; Skophammer et al. 2007). Comparative methods using cellular and molecular biology also gave strong evidence for the bacterial rooting system, especially for Gram-negative bacteria (Cavalier-Smith 2002; Skophammer et al. 2007). The implication followed that the Bacteria branched first from the universal tree with Archaea and Eukarya as sister groups.

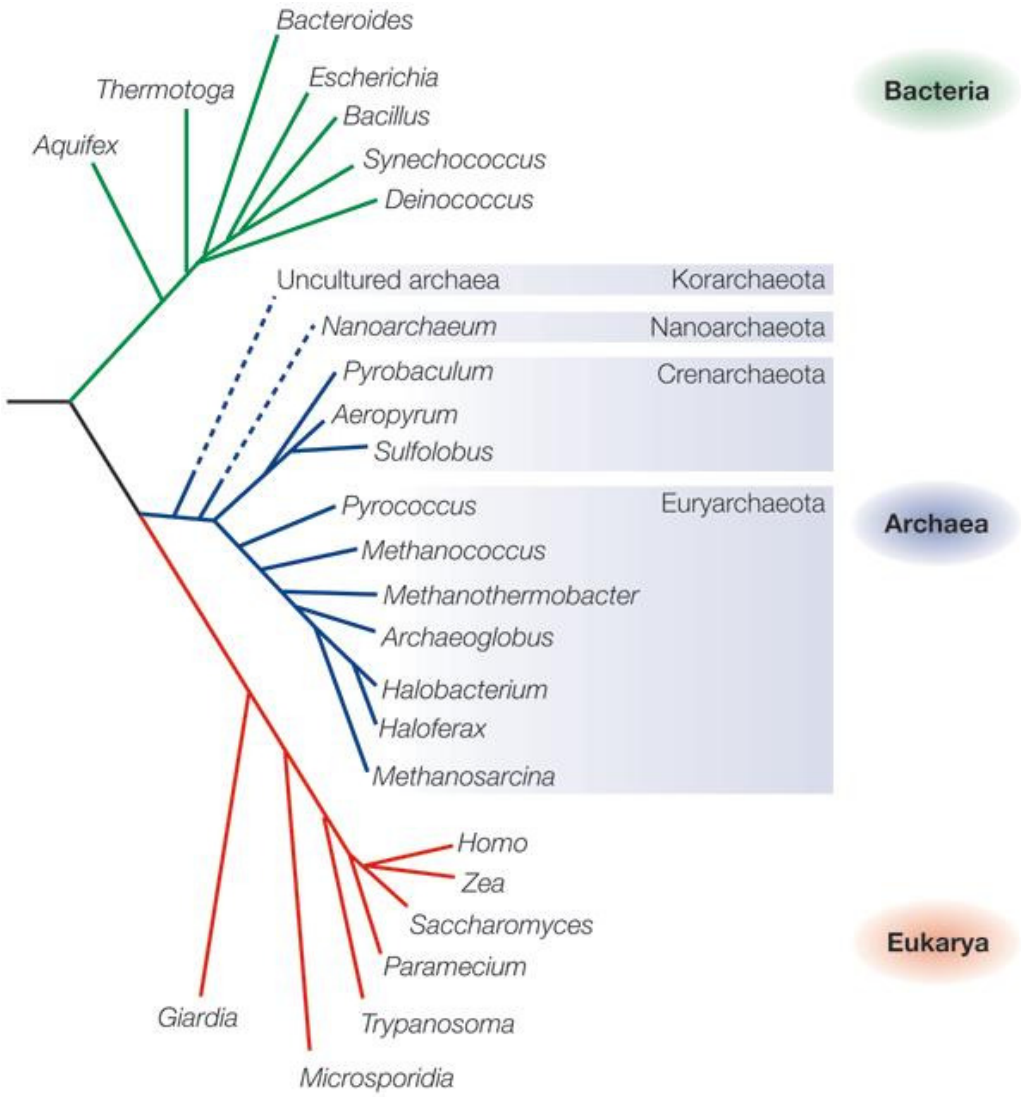


Figure 1.1: Universal Tree of Life
This phylogenetic tree was constructed using ribosomal RNA sequences to illustrate the tripartite view of evolutionary divergence. A novel species of the Korachaeota has since been characterized. Reproduced from Allers and Mevarech (2005).

Archaea have characteristics that are unique compared to the other two domains; this includes the different lipid structure within the membranes of the Archaea such as the presence of isoprenol ether lipids instead of ester linkages found in Bacteria and Eukarya (De Rosa et al. 1986). Some of the lipid biosynthetic enzymes are also different in Archaea (Chen and Poulter 1993). Nevertheless, Archaea share many features of both Bacteria and Eukarya. For instance, Archaea and Bacteria are similar in physical morphologies such as cell size, lack of nuclear membrane and organelles, and the presence of a large circular main chromosome sometimes accompanied by smaller circular DNA plasmids (Keeling et al. 1994). In terms of

metabolic diversity, most of the metabolic pathways discovered in Archaea, except for methanogenesis, are also found in Bacteria (Huber et al. 2000). Many operational proteins, such as those involved in primary and secondary metabolism, membrane receptors and transporters and cell division are also found to be very similar between the two prokaryotic domains (Koonin et al. 1997; Jain et al. 1999). In contrast, the Archaea genomes have shown to be very homologous to that of the Eukarya in the number of similar informational proteins encoded. This is most evident in the discovery that Archaea encode homologues of nearly all eukaryal DNA replication proteins, compared to just one homologue to the bacterial DNA replication proteins (Edgell and Doolittle 1997; Forterre 1999; Leipe et al. 1999). As it is often considered that informational proteins form the basis of any organism's genome, they are less likely to be subjected to lateral gene transfer (LGT) and therefore, more representative of an organism's actual history. This has been documented in some studies that demonstrated the lack of LGT of informational proteins between Archaea and Bacteria (Brochier et al. 2000; Brochier et al. 2002).

Many archaea thrive in harsh environments, a feature that could be indicative of their ancestral evolutionary divergence. According to 16S rRNA sequence comparisons, the archaeal domain is currently organized into two major lineages: the Euryarchaeota (from the Greek "*euryos*" meaning diversity) and the Crenarchaeota (from the Greek "*crenos*" meaning spring or origin) (Woese et al. 1990). The Euryarchaeota included a mixture of methanogens, extreme halophiles, thermoacidophiles and a few hyperthermophiles; in contrast, the Crenarchaeota only included hyperthermophiles. This division of the Archaea was widely accepted due to the early discovery of fundamental differences in SSU rRNA oligonucleotide catalogues and RNA polymerase structures between species of the two phyla (Prangishvilli et al. 1982; Woese et al. 1984). Nevertheless, the discovery of 19 additional archaeal sequences from a hot spring in Yellowstone National Park has since expanded the apparent diversity of Archaea. On the basis of PCR amplifications of 16S rRNA genes from environmental DNA, Korarchaeota was postulated as a possible third phylum within the Archaeal kingdom due to their unusually deep branching ribosomal RNA sequences in phylogenetic trees (Barns et al. 1996). The first member of the Korarachaeota with a proposed name, *Candidatus Koarachaeum cryptofilum*, was recently described to exhibit ultrathin filamentous

morphology (Elkins et al. 2008). However, the cultivation of any members of this proposed phylum in pure culture remains elusive. Furthermore, the cultivation of a novel nanosized hyperthermophilic archaeon from a submarine hot vent has led to the representation of a new phylum, Nanoarchaeota (Huber et al. 2002; Huber et al. 2003). The named species, *Nanoarchaeum equitans*, grows anaerobically in nature and symbiotically with the crenarchaeon, *Ignicoccus hospitalis*, a new member of the genus *Ignicoccus*. Genomic sequences of both species have recently been published, allowing for the first time a comprehensive genomic analysis of this unique Archaeal association (Waters et al. 2003; Jahn et al. 2008; Podar et al. 2008; Forterre et al. 2009). Lastly, *Thaumarchaeota* was proposed as a novel phylum based on the first genome sequence of a mesophilic crenarchaeote, *Cenarchaeum symbiosum*. It was shown to be different from the hyperthermophilic Crenarchaeota and formed deep branching before the speciation of the Crenarchaeota and the Euryarchaeota (Brochier-Armanet et al. 2008).

The use of the genome-based approach to phylogenetics (phylogenomics) has greatly expanded in the past few years. This arose in part due to the increase in the number of genomes sequenced, and as a way to compensate for the sampling effects problems faced with using classical phylogenetic inference methods (Rokas et al. 2003; Soltis et al. 2004; Delsuc et al. 2005). This genome-based approach has been used to identify archaeal-specific proteins that provide a novel insight into the evolutionary relationships among different groups within Archaea, specifically regarding the origin of methanogenesis (Sicheritz-Ponten and Andersson 2001; Gao and Gupta 2007). It has also highlighted a core of vertically inherited genes within Archaea, such as those coding for the flagellum, to provide a globally well-reserved picture of Archaeal evolution (Gribaldo and Brochier-Armanet 2006; Desmond et al. 2007). Interestingly, it has also demonstrated the occurrences of rare horizontal gene transfer within lineages (Desmond et al. 2007; Forterre and Gadelle 2009). With the growing availability of the completed genome sequences from cultivated archaeal species as listed in a recent publication (Allers and Mevarech 2005), a phylogenomics approach provides a better understanding of the molecular evolutionary processes and serves as an important framework for tracking genomes of uncultivated archaea in complex communities.

1.2 Extremophiles

Extremophiles are microorganisms that have evolved to exist in a number of different extreme environments. Macelroy first coined the term "extremophiles" more than three decades ago (Macelroy 1974). Since then, the progression of the study of these organisms has been steady; in 1996, the First International Congress on Extremophiles was convened in Portugal. A year later, the scientific journal, Extremophiles, was published. The emerging interest in the field was largely due to the isolation and cultivation of novel organisms from environments that were once considered incompatible with biological material and impossible to sustain life. This new-found appreciation for the diversity of microorganisms brings with it challenging concepts and ideas regarding the origin and evolution of life on Earth. The discovery of novel species and their cellular components also struck an interest in the biotechnology industry, offering new and potential applications in various fields.

The majority of extremophiles belong to the phylum Archaea (Woese et al. 1990); however, extremophilic bacteria have also been described. Extremophiles include organisms adapted to high temperature of 55° to 121°C (thermophiles), low temperature of -2° to 20°C (psychrophiles), high acidity with a pH below 4 (acidophiles), high alkalinity with a pH above 8 (alkaphiles) and high salinity between 2 and 5M NaCl (halophiles). Other classes exist, such as those adapted to high pressure (piezophiles), high levels of metal concentration (metallophiles) and high levels of radiation (radiophiles) (Madigan and Marrs 1997; Rothschild and Mancinelli 2001). Some of these extreme environments occupied include hot springs and geysers, deep sea environments, hypersaline environments, evaporites, desserts and high altitude glaciers. A few extremophiles also thrive in unusual conditions such as rocks buried deep beneath the surface of the Earth, areas of extreme dryness, low water and low oxygen supply. In addition, extremophiles can be found in environments with a combination of extreme conditions, such as thermoacidophiles, capable of surviving both high temperature and low pH.

The advances made in the field of genomics and proteomics have vastly increased the number of identified gene products based on comparison with gene sequences of

proteins of known functions from other organisms. These enzymes, or extremozymes, produced from extremophilic species offer several potential uses in the biotechnology sector (Hough and Danson 1999). Specific archaeal metabolites have also been purified and characterized, some of which have potential industrial uses. As the majority of enzymes used to date have been isolated from mesophilic organisms, their applications are restricted by their limited stability to extreme conditions. The harsh conditions of many industrial processes also call for biocatalysts that can withstand these conditions. Thus, the availability of a growing number of completed genomes, along with improved annotation techniques, will provide means of uncovering novel, archaeal extremozymes. With their extreme stability and activity, these enzymes offer new opportunities for biocatalysis and biotransformation. Some of these potentially useful extremozymes include cellulases, amylases, xylanases, proteases, pectinases, keratinases, lipases, esterases, catalases, peroxidases and phytases (Eichler 2001; Gomes and Steiner 2004). The improvements in the techniques for cultivation in pure cultures and large scale production for industrial uses, along with better molecular biology tools for the cloning and expression of the genes in mesophilic hosts, will certainly increase the biocatalytic applications of extremozymes. The unraveling of new metabolic pathways in the Archaea may further increase the diversity of the enzymes available, thus contributing to the pool of extremozymes available to meet the growing industrial and biotechnological demand and interest of these proteins.

1.3 Halophilic archaea

1.3.1 Taxonomy of halophilic archaea

The pioneering work of the late Professor Benjamin Elazari-Volcanii on the studies of the microflora of the Dead Sea revolutionized the way microbiologists viewed the Dead Sea as a completely sterile environment in the early 1900s. It was found to be a breeding ground for numerous prokaryotic and unicellular eukaryotic species, including the most abundant halobacteria whose cell density can reach 10^7 cells per ml. Traditionally, the term “halobacteria” referred to all extremely halophilic microorganisms, including both halophilic bacteria and halophilic archaea. In 1974, Gibbons established the family *Halobacteriaceae* to accommodate the genera, *Halobacterium* and *Halococcus* (Gibbons 1974). In the modernized version of

taxonomy and terminology, those that belong in the archaeal domain fell into a single order, *Halobacteriales*, and family *Halobacteriaceae* (Grant and Larsen 1990). Haloarchaea is the common named applied to members of this class, and consists of the extremely halophilic archaea. The Halobacteriaceae family now includes 27 genera as of September 2008, where the 3-letter abbreviation system is currently in use (Oren and Ventosa 2000; Oren et al. 2009). A most up-to-date account of the taxa including some relevant references for each genus can be found online (http://www.the-icsp.org/taxa/halobacterlist.htm). The 27th genus, which is not shown on the page, is derived from the discovery of a novel halophilic archaeon, *Halosarcina pallida*, where comparisons of various nature suggests that it represents a member of a new genus (Savage et al. 2008).

Most genera are found in salterns and salt lakes, where the pH, salinity, temperature and the concentration of various ions vary from high to low for these environments, contributing to the different requirements of the halobacteria (Oren 2002). For instance, *Natrinema pellirubrum* was isolated from an unused saltern with a low salt concentration of 0.2M NaCl, while the majority of the halobacteriaceae species require a minimum concentration of 1.8M NaCl in their saline habitats (McGenity et al. 1998). In terms of ion requirement, the haloalkaphilic archaeon, *Halalkalicoccus tibetensis*, flourishes in the alkaline chloride-sulfate salt lake of Zabuye with a pH optimum of 9.5, where Mg ions are not required for growth (Xue et al. 2005). On the other hand, *Halogeometricum borinquense*, an isolate from a solar saltern in Puerto Rico, requires 40-80mM of Mg for maintaining its growth (Montalvo-Rodriguez et al. 1998). Halobacteria have also been isolated from habitats other than salterns and salt lakes. For example, *Natrialba asiatica* was discovered from a survey of dry beach sands with granular salts attached (Kamekura et al. 1997). Meanwhile, crystals of *Halosimplex carlsbadense* were removed from an unsterilized rock salt taken from a Permian-aged Salado Formation (Vreeland et al. 2002).

With the increase in the number of sites examined comes the discovery of new *Halobacteriaceae* genera and species as well as information on their diversity. *Halorubrum cibi* sp. nov., isolated from salt-fermented seafood, is an example of a newly characterized strain with identified genotypic and phenotypic differences from

the other *Halorubrum* species (Roh and Bae 2009). Similarly, novel species of the *Natrinema* genus, *Natrinema gari* sp. nov., and of the *Halobacterium* genus, *Halobacterium piscisalsi* sp. nov., were discovered from salt-fermented fish products (Tapingkae et al. 2008; Yachai et al. 2008). Some novel, unclassified species that bear possible lineages to members of the *Halobacteriaceae* family have also been documented from sites such as the former Saltern Lake Tescoco in Mexico (Valenzuela-Encinas et al. 2008). More important, however, is the correlation between increased microbiodiversity in hypersaline environments and the constant changing condition of these environments as observed in the Great Salt Plains of Oklahoma and the hypersaline lakes in Egypt (Mesbah et al. 2007; Caton et al. 2009).

1.3.2 Physiology of halophilic archaea

Halophilic and halotolerant microorganisms exist in all three domains of life (Oren 2002). Species of the archaeal domain can be distinguished from the other two domains by their vastly different membrane lipid structure and composition. Archaeal lipids are characterized by ether linkages and isoprenoid chains, which is in constract to ester linkages and straight fatty-acid chains of bacteria and eukarya (De Rosa et al. 1986; Sprott 1992; Kates 1993). In terms of composition, the membranes of extremely halophilic archaea contain several unique characteristics that vary little between genera, such as the absence of phospholipids with ethanolamine, inositol and serine head groups (Kamekura and Seno 1990). Some of the lipid biosynthetic enzymes are also different in Archaea. For example, distinct farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP) enzymes are found in eukaryotes, whereas a single bifunctional enzyme exists in archaea (Chen and Poulter 1993). A further distinction can be made by their adaptation strategies to environments of high salinity. To balance the high external osmotic pressure, bacteria either obtain organic solutes through uptake in the environment, or by the synthesis of compatible osmolytes, *de novo* or from storage material (Oren 1999). These compatible solutes or osmolytes usually consist of amino acids and polyols, such as glycine betaine, ectoine, sucrose, trehalose and glycerol, all of which do not disrupt metabolic processes and have no net charge at physiological pH (DasSarma and Aurora 2002). Extreme halophilic archaea, however, maintain osmotic balance by the accumulation

of KCl equal to the external concentration of NaCl through osmoregulation, where their membranes are adapted to a low proton and sodium permeability at high salt concentrations (van de Vossenberg et al. 1999; Margesin and Schinner 2001).

Hypersaline environments are generally defined as those with a salt concentration in excess of sea water. These include naturally-occuring hypersaline lakes such the Great Salt Lake in Utah, USA, and the Dead Sea in Israel, Jordan and the West Bank, two of the largest and best-studied habitats for halophilic microorganisms. Near the coastal areas, many small, hypersaline evaporations are found; examples of these include the Ras Muhammad Pool near the Red Sea coast, Guerrero Negro on the Baja Californiа coast, and the Deep Lake of Antarctica. Alkaline hypersaline soda brines can also be found in Egypt, Kenya and the United States. Lastly, artificial solar salterns have been constructed to produce sea salts, where a high cell density (10^7 ml^{-1}) of *Halobacterium* species is commonly observed (DasSarma and Aurora 2002).

Most halophilic archaea are strict aerobes or facultatively anaerobic, chemoorganotrophs, found ubiquitously in hypersaline brines (Dundas 1977; Hartmann et al. 1980; Rodriguez-Valera 1988). Reports of anaerobic growth have also been documented for some halophilic archaea (Hartmann et al. 1980), either using nitrate as an alternative electron acceptor (Mancinelli and Hochstein 1986; Oren and Truper 1990; Wanner and Soppa 1999) or by fermentative L-arginine utilization (Oren and Truper 1990; Ruepp and Soppa 1996). Other sources of electron acceptors include dimethylsulfoxide, trimethylamine N-oxide and fumarate. Most halophilic archaea grow at near neutral pH between 7.2 and 7.5, whereas the alkalophilic haloarchaea grow optimally between pH 8.5 and 10. Halophilic microorganisms are classified according to their response to salt; extreme halophilic archaea grow best at a salt concentration of 2.5-5.2M. It was found that when suspended in a solution of less than 1-2M salt, cells were irreversibly damaged, and many species started to lyse (Oren 2006).

The temperature optima for most strains of the halophilic archaea occur between 30°C and 50°C (Oren 2001). Interestingly, the Antarctic isolate, *Halorubrum*

lacusprofundi, grows optimally between 31° C and 37° C, whereas for eight months the temperature of the Deep Lake is less than 0°C (Franzmann et al. 1988). A recent study of 14 genera of *Halobacteriaceae* revealed that the temperature optima for maximum growth rates occurred between 49°C and 58°C; this is not surprising given that many haloarchaea are exposed to high temperatures in their natural habitats (Robinson et al. 2005). Oxygen solubility decreases with an increasing level of salt, especially at raised temperatures. To overcome the problem with limited oxygen for respiration, halophlic archaea have adapted to slow growth rates, coupled with an efficient terminal oxidase for catalyzing reactions with molecular O_2 (Dyall-Smith 2008).

A special feature of the halophiles is their purple membrane, a specialized region of the cell membrane which contains bacteriorhodopsin (bR). Bacteriorhodopsin is a protein (bacterioopsin)-chromophore (retinal) complex, which serves as a light-driven proton pump first discovered in *Halobacterium salinarum* (Oesterhelt and Stoeckenius 1971; Oesterhelt and Stoeckenius 1973). This proton pump is induced by high light intensity and low oxygen tension; a flash of light was shown to induce a transient shift of its absorption maximum, accompanied by release and uptake of protons as well as reduced O_2 consumption. When exposed to light and devoid of any energy sources, starved and anaerobic cells that contain purple membrane are able to generate and maintain a proton gradient across their cell membranes (Oesterhelt and Stoeckenius 1973). This generated membrane potential can be used to drive ATP synthesis and to support a period of phototrophic growth.

Another interesting feature of the halophilic archaea is the production of carotenoid pigments which gives them a characteristic orange to red colour. These pigments were suggested to stimulate an active photorepair system for the repair of UV-induced thymine dimers (Dundas and Larsen 1963). Some halophilic archaea also have buoyant gas vesicles that enable the cells to float to the more oxygenated surface layers as these microorganisms are primarily aerobes that occupy concentrated brines where oxygen solubility is low (Walsby 1994; DasSarma and DasSarma 1997). This also allows for the purple membrane-mediated

photophosphorylation to occur with the availability of light (Clark and Walsby 1978).

Halophilic archaea exist as cocci, rods, flat discs (*Halococcus*), squares (*Haloferax*), rectangles (*Halobacterium*) and triangles (*Haloarcula*) (Grant et al. 2001). The various cell shapes are as shown in Figure 1.2. With the exception of the halocci, these species lack rigid cell walls and instead contain a single layer of glycoprotein called the surface layer, or S-layer. This was demonstrated as early as 1956 for *Halobacterium salinarum*, where subsequent studies using other archaeal species have also confirmed such observations (Houwink 1956; Lechner and Sumper 1987; Lechner and Wieland 1989; Messner 1997). All known S-layer glycans, with one exception, consist of short, linear chains of up to ten sugar residues. The only long glycan, a heavily sulfated heteropolysaccharide, was found in *Halobacterium salinarum*. These glycan chains are linked to the S-layer polypeptide by N-glycosidic linkages. The subunits of the S-layer are held together by divalent cations Mg^{2+}, and can be completely removed by treating the cells with EDTA (Messner 1997). The biological function of the glycoprotein in haloarchaea is one related to maintaining cell structure; it has been demonstrated that upon removal of the N-glycosidic linkages, the rod shaped *Halobacterium salinarum* changed its structure to round spheres and that the process is irreversible (Mescher and Strominger 1976).

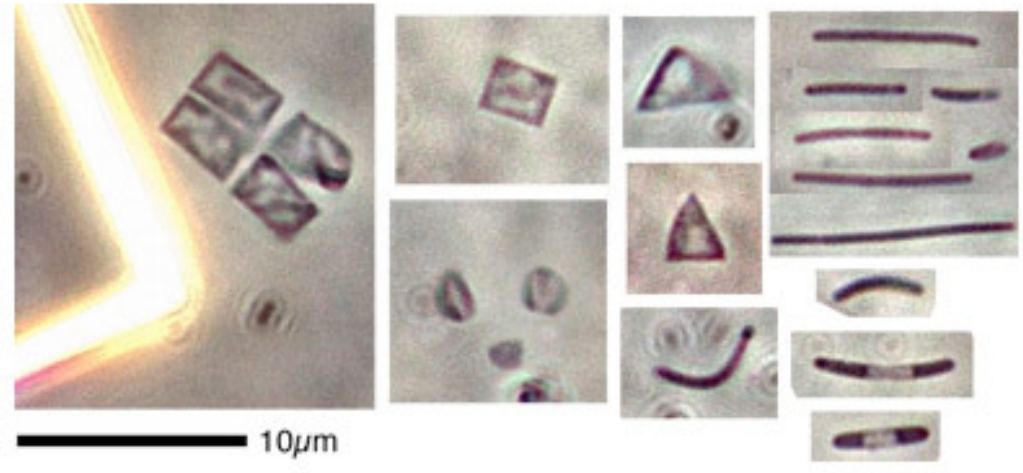


Figure 1.2: Various cell morphologies of halophilic archaea found in salt lake water. From left to right: square, discs, triangles and rod shaped halophiles. Reproduced from Dyall-Smith (2008).

In 1980, a square halophilic archaeon, Walsby, was described in water samples taken from a hypersaline pool near the Red Sea (Walsby 1980). These cells were extremely thin squares which contained gas vesicles, and formed extended sheets. The cells were of special interest due to their unique shape and their abundance in hypersaline waters. It was not until recently that the first isolates subsequently named *Haloquadratum walsbyi*, were obtained in pure culture by two independent groups (Bolhuis et al. 2004; Burns et al. 2004; Walsby 2005). In 2007, two different isolates were determined, where phylogenetic and phenotypic characteristics indicated that they belong to a novel species and a new genus within the *Halobacteriaceae* family (Burns et al. 2007). Wasby's square halophilic archaeon is shown in Figure 1.3.

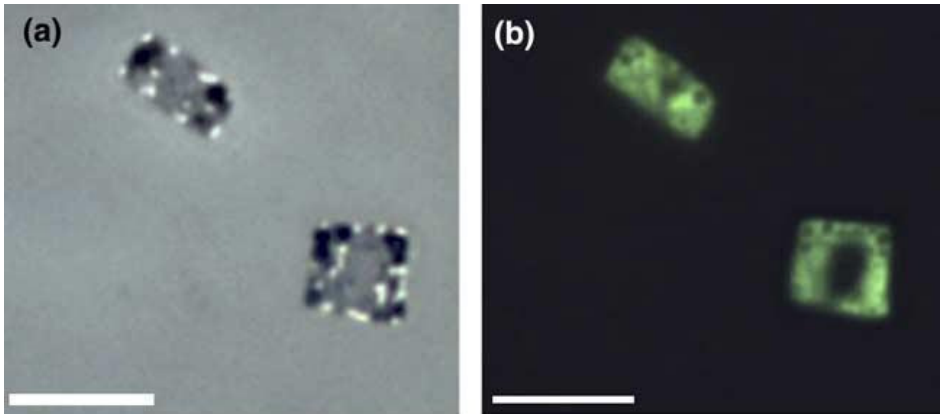


Figure 1.3: Waslby's square halophilic archaeon (a): cells examined by phase contrast microscopy; (b): the same field viewed by epifluorescence microscopy after acridine orange staining. Reproduced from Burns *et al*. (2004).

1.3.3 Genomics and functional genomics of halophilic archaea

The genome of *Halobacterium* sp. NRC-1 was published in 2000 (Ng et al. 2000); subsequently, the genome project for a related type strain *Halobacterium salinarum* was also completed (Pfeiffer et al. 2008). Comparison of the two genomes revealed that while the main chromosome is virtually identical, the many smaller replicons ared very different. In 2004, the genome sequence of *Haloarcula marismortui* was published (Baliga et al. 2004) and the genome sequence of *Haloferax volcanii* can currently be found on the UCSC website (http://archaea.ucsc.edu/cgi-bin/hgGateway?db=haloVolc1). Furthermore, sequences of over 1000 random genomic clones of *Ha. marismortui* and three additional species, *Natrialba asiatica, Halorubrum lacusprofundi,* and *Halobaculum gomorrense*, have been reported (Goo

et al. 2004). The sequences of both *Natronomonas pharaonis* (Falb et al. 2005) and *Haloquadratum waslbyi* (Bolhuis et al. 2006) have since been published.

The common feature shared amongst all haloarchaea includes a major chromosome with a high GC rich content of 63% up to 70%. Some of the smaller replicons of *Halobacterium haloarcula* and *Haloferax volcanii* contain essential genes; thus, it is possible that the occurrence of additional chromosomes may be a common phenomenon. All proteins have a very low isoelectric point, an adaptation to the high salt concentration of the cytoplasm. It is also apparent that some genes are present in multiple copies in haloarchaea, whereas in other archaea they are single-copy genes. Until very recently, prokaryotic chromosomes are believed to contain a single origin of replication; therefore, the discovery that *Sulfolobus solfataricus* contains three and *Aeropyrum pernix* contains two origins of replications was surprising (Robinson et al. 2004; Robinson and Bell 2007). For *Halobacterium salinarum*, it has been proposed that two replication origins exist; however, *in vivo* analysis of three putative replication origins could only verify the *in vivo* function of one chromosomal site (Berquist and DasSarma 2003). On the other hand, the *in vivo* activity of three replication origins on the major chromosome has been demonstrated in *Haloferax volcanii*, with two origins on an additional replicon (Norais et al. 2007). A further commonality lies in the existence of transducers found in all species, an indication that halophiles are capable of sensing and reacting to a wide variety of environmental signals consistent with their high metabolic versatility (Soppa 2006).

The use of both *Haloferax volcanii* and *Halobacterium salinarum* allows for the transfer of biological processes from one to another, leading to the identification of all essential elements. As an example, using complementation studies, the complete gas vesicle biosynthesis pathway was successfully transferred from *Halobacterium* to *Haloferax*, which is naturally devoid of gas vesicle genes (Offner and Pfeifer 1995). Recent reports include characterization of a regulatory system for gas vesicle formation in *Halobacterium mediterranei* (Zimmermann and Pfeifer 2003), and of the translational initiation at gvp transcripts (Sartorius-Neef and Pfeifer 2004; Scheuch et al. 2008).

The development of DNA microarrays for both *Haloferax volcanii* and *Halobacterium salinarum* has been extremely useful in characterizing the regulation of metabolic pathways and protein complexes in response to changes in environmental conditions, allowing for the parallel investigation of all genes involved. These have been applied to studies in the genome-wide characterization of energy metabolism regulation, stress responses and differences between wild types and mutants (Soppa et al. 2008). Energy metabolism studies include those of phototropic and aerobic growth, (Baliga et al. 2002; Zaigler et al. 2003; Twellmeyer et al. 2007), anaerobic respiration (Muller and DasSarma 2005), characterization of cellular responses to sudden anaerobiosis (Schmid et al. 2007), and the elucidation of a pentose degradation pathway (Schönheit and Soppa, unpublished data). The characterization of global stress responses include the exposure to UV irradiation (Baliga et al. 2004; McCready et al. 2005), to gamma radiation (Whitehead et al. 2006), to transition metals (Kaur et al. 2006), to various temperatures and salinities (Coker et al. 2007) and to phosphate limitation (Wende et al. 2009). The mutants versus wild type studies include radioresistant mutants generated by random mutagenesis (DeVeaux et al. 2007) and the construction of isogenic *in frame* deletion mutants (Coker and DasSarma 2007; van Ooyen and Soppa 2007; Dambeck and Soppa 2008; Tarasov et al. 2008).

Proteomic analyses are also available for halophilic archaea, notably *Halobacterium salinarum*, *Haloferax volcanii* and *Natronomas pharaonis*. Compared to transcriptome analysis, however, proteomic analyses are much less comprehensive, meaning that several approaches have to be combined to produce a high coverage of the theoretical proteome. With proteomic studies, the major aims are to establish a sound methodology and to identify all the proteins produced by the respective species; this forms the basis for half of the publications in this area. For *Halobacterium salinarum*, the cytosolic proteome (Tebbe et al. 2005), membrane proteome (Klein et al. 2005) and low molecular weight proteome (Klein et al. 2007) as well as a large scale study of N-terminal peptides (Aivaliotis et al. 2007) have been established, making it a reference species for proteomics research. Recently, the approach has shifted to studying gene expression at the protein level using qualitative comparisons under two different conditions. This includes identification of proteins induced by heat shock (Shukla 2006), salinity-mediated regulation of the

stress response protein (Bidle et al. 2008), membrane protein during phototrophic and aerobic growth (Bisle et al. 2006), the induction of proteins in presence and absence of oxygen or isopropyl alcohol (Ha et al. 2007; Schmid et al. 2007), characterization of proteosome function (Kirkland et al. 2007; Kirkland et al. 2008) and a large-scale study of posttranslational modification of proteins (Falb et al. 2006).

1.4 *Haloferax volcanii*

Haloferax volcani (*Hfx. volcanii*), an obligate halophile, was first isolated from the bottom sediment of the Dead Sea in 1975 (Mullakhanbhai and Larsen 1975). It belongs to the family Halobacteriaceae, and was previously described as *Halobacterium volcanii* sp. nov. It is disk-shaped, and in the presence of salt it becomes cup-shaped. *Hfx. volcanii* is a chemoorganotroph, and grows aerobically in both complex and minimal media (Mevarech and Werczberger 1985). It has a salt requirement between 1.5M and 2.5M; whilst it will grow at 37° C, the optimal temperature is 42° C. A recent survey of growth kinetics of various halophilic archaea shows that *Hfx. volcanii* can withstand a maximum temperature of 49° C, where growth was unsustainable above this temperature (Robinson et al. 2005). The completed genome project indicates that the *Hfx. volcanii DS2* genome is 4.01 million base pairs long, and contains approximately 4209 predicted genes. The genome has a multireplicon structure consisting of a main chromosome of 2.9 Mb and four smaller replicons: pHV1, pHV2, pHV3 and pHV4 (Charlebois et al. 1991).

The continuing interest in halophilic organisms has led to the development of biochemical, genetic, genomic and cell biology tools for use in the presence of the physiological high salt concentrations of the haloarchaeal cytoplasms. As such, halophilic archaea have developed into excellent model organisms for the study of archaeal biology. *Hfx. volcanii*, especially, has some of the best genetic tools available, including the first archaeal transformation system (Cline et al. 1989). More recently, an efficient gene-knockout system called the "pop-in-pop-out method" for making in-frame deletions of genes and to exchange chromosomal genes with mutant variants was developed (Bitan-Banin et al. 2003). A number of plasmids have been constructed containing antibiotic resistance genes for selection

(Lam and Doolittle 1989; Holmes and Dyall-Smith 1990; Holmes et al. 1994). These plasmids have been further refined to include integrating and shuttle vectors based on 4 different selection markers: *pyrE2*, *leuB*, *trpA*, and *hdrB* (Allers et al. 2004). This method allows for the introduction of recombinant DNA fragments into the chromosome and the deletion of the native copy by both positive and negative selection using designed host strains and plasmids. A complementation system has been reported based on designed mutants with an inactivated essential function. This leads to the identification of the function of novel, hypothetical proteins able to complement the missing function (Levin et al. 2004). A plasmid containing a tightly controlled promoter of the tryptophanase gene was recently characterized for *Hfx. volcanii*; this is useful in studies of complementation and depletion of essential gene products (Large et al. 2007). A reporter system using the β-galactosidase gene has also been described previously for both *Hfx. volcanii* and *Halobacterium salinarum* (Holmes and Dyall-Smith 2000; Patenge et al. 2000). More recently, *Hfx. volcanii* has been used to search for the additional existence of sRNAs (small non-coding RNAs) in archaeal genomes using bioinformatics and Rnomics approaches and the availability of both transformation and gene knockout system (Soppa et al. 2009; Straub et al. 2009).

1.5 Halophilic enzymes

Due to the high intracellular salt concentration accumulated within halophilic archaea, all cellular components must be adapted to function under such condition. Halophilic enzymes are no exception; the interest in the salt dependence, stability and catalytic properties of these enzymes has been a well sought-after area of research. Halophilic enzymes display a remarkable instability in low solvent salt concentrations, and are usually only stable in solvents of very high salt concentrations, under which most proteins from the non-halophilic organisms are likely to unfold or precipitate (Eisenberg and Wachtel 1987). Consequently, heterologous expression of halophilic proteins in *Escherichia coli*, such as the *Hfx. volcanii* citrate synthase, often results in soluble but inactive enzymes that require reactivation by the addition of high salt. More likely, however, is the formation of insoluble aggregates that require re-solubilization through unfolding with a chaotropic agent, followed by refolding in high salt concentrations (Connaris et al.

1999). Examples of this include the *Hfx. volcanii* dihydrolipoamide dehydrogenase, the NADP-dependent isocitrate dehydrogenase and the *Hfx. mediterranei* NAD-glutamate dehydrogenase (Connaris et al. 1999; Camacho et al. 2002; Diaz et al. 2006).

Studies on the amino acid composition of halophilic enzymes have shown a higher proportion of acidic over basic residues (Lanyi 1974; Rao and Argos 1981; Benachenhou and Baldacci 1991). A statistical analysis of over 26 halophilic proteins has confirmed the acidic nature of these proteins, showing a reduction of lysine residues and an increase in small, hydrophobic ones, i.e. glycine, valine, and alanine (Madern et al. 1995). The relationship between acidic residues and salt binding was first suggested using a stabilization model of tetrameric malate dehydrogenase (MalDH) from *Haloarcula marismortui* (Zaccai et al. 1989; Bonnete et al. 1993). It has been proposed that the excess of acidic residues, predominantly located on the enzyme surface, leads to the formation of a hydration shell that protects the enzyme from aggregation in its highly saline environment. This is due to the superior water-binding capabilities of these acid groups, allowing the protein to compete with salt ions for free water at high salt concentrations (Frolow et al. 1996). This has been demonstrated from various halophilic enzymes such as malate dehydrogenase, 2Fe-2S ferrodoxin, dihydrofolate reductase, catalase peroxidase and a novel dodecameric flavin binding protein (Dym et al. 1995; Frolow et al. 1996; Pieper et al. 1998; Yamada et al. 2002; Bieger et al. 2003).

A second characteristic of the surface of halophilic enzymes that may contribute to their resistance to high salt concentration is the display of a significant reduction of hydrophobicity and accumulation of negative charges. This view was first demonstrated using a modeled structure of *Hfx. volcanii* dihydrofolate reductase, which showed clustering of acidic residues on the surface of the enzyme (Bohm and Jaenicke 1994a; Bohm and Jaenicke 1994b). It was proposed that the electrostatic repulsion of the clusters of acidic residues can only be tolerated at high salt concentrations. A decrease in salt concentration leads to the destabilization of the protein until it is completely unfolded. This, however, can be offset by a decrease in the pH of the solution, leading to the protonation of the acidic groups and thus reducing the unfavourable electrostatic interactions. A theoretical approach to the

study of this pH-dependence of stability has recently been reported (Elcock and McCammon 1998).

1.6 Tolerance to UV radiation

In their natural habitats, halophilic archaea are exposed to intense sunlight where the cells receive substantial doses of ultraviolet (UV) irradiation. Ultraviolet radiation induces the formation of thymine dimers within the DNA, blocking replication and transcription, thus affecting the viability of the organism (Friedberg et al. 1995). Halophilic archaea, however, are superbly adapted to such UV intense environments, leading to the suggestion that numerous repair pathways must be in place. The use of *Halobacterium* as a model species has paved the way to the elucidation of some evolutionary conserved repair systems that are present in all three domains, as well as specialized mechanisms that allow the halophilic archaea to flourish under constant UV irradiation.

UV irradiation causes inducible responses in all organisms known to date. This results in an upregulation of gene expression to allow for the repair of DNA lesions. The changes in gene expression in reponse to UV damage due to UV and other environmental stresses have been collectively termed the SOS response. In bacteria, the RecA and LexA gene products coordinate the upregulation of more than 20 genes, including *recA*, *uvrABC*, *sulA*, *umuCD*, *lexA*, and *ruvAb* (Courcelle et al. 2001; Quillardet et al. 2003). In eukaryotes, various genes are also upreguated in reponse to UV damage; however, no equivalent bacterial SOS response has been detected (Cline and Hanawalt 2003). While an earlier report indicated the possible presence of an SOS-like response in the hyperthermophilic archaeon Pyrococcus (Bouyoub et al. 1995), it has since been recognized that the classical bacterial SOS response is absent in archaea . This is demonstrated by Baliga et al., where, with the exception of a moderate increase in *radA1* mRNA expression, no coordinated change in the required repair genes was observed with UV exposure in *Halobacterium* sp. NRC-1 (Baliga et al. 2004). A study by McCready and colleagues also showed that neither eukaryotic nor prokaryotic excision repair genes were induced post UV and there was a lack of an SOS-like reponse. However, *radA1* was shown to be highly induced along with numerous other genes involved in homologous recombination

(McCready et al. 2005). Furthermore, a core set of genes, including *RadA1*, was shown to be consistently up-regulated after exposure to biologically relevant low UV-B doses of UV in *Halobacterium* sp. NRC-1. Results indicated that in 4 of the up-regulated genes, an 11-base pair motif upstream of the TATA box regulator was found, suggesting its role as a binding site for a transcriptional regulator involved in their response to UV damage (Boubriak et al. 2008).

In *S. solfataricus*, genome-wide transcriptional studies have indicated that no genes involved in the direct repair of DNA damage were upregulated strongly, again suggesting that the typical SOS response is not found in Archaea (Frols et al. 2009). In another microarray study with both *S. solfataricus* and *S. acidocaldarius*, while there was a clear transcriptional response to DNA damage causing the down-regulation of DNA replication machinery and an up-regulation of the biosynthetic enzyme for beta carotene production, no induction of DNA repair proteins were observed (Gotz et al. 2007). However, conficting results were observed in Salerno et al., where exposure to UV induced transcriptional regulation of potentially NER associated genes (Salerno et al. 2003) and recently an active NER system has been demonstrated (Dorazi et al. 2007; Romano et al. 2007). Nucleotide excision repair (NER) is a light independent repair pathway common to both bacteria and eukaryotes for the removal of UV-induced lesions. In addition, bacterial NER gene homologs have also been identified in both *M. thermautotrophicus* and *Halobacterium* sp. NRC-1 where the requirement for the repair genes in removing UV induced lesions has been demonstrated (Ogrunc et al. 1998; Crowley et al. 2006).

Of all the genes potentially required for DNA repair, only those belonging to the *rad50/mer11* operon showed slight upregulation due to UV in *S. solfataricus* (Frols et al. 2007; Gotz et al. 2007). These are homologous to eukaryotic and bacterial proteins involved in double stranded break (DSB) repair via homologous recombination. The suggestion that homologous recombination plays an essential role in the UV-induced response has been demonstrated in *Halobacterium* sp. NRC-1, where, upon UV irradiation, the *radA1* transcript was induced to a high level. RadA1, which has been shown to cause severe UV sensitivity in *Hfx. volcanii* when it is knocked out (Woods and Dyall-Smith 1997), is the archaeal homolog of

RecA/Rad51 that catalyzes strand invasion and exchange during homologous recombination (McCready et al. 2005).

Ultraviolet radiation induces the formation of dimers, or kinks, within DNA as illustrated in Figure 1.4. Of the two major types of lesions, cyclobutane pyrimidine dimers (Pyr<>Pyr), or CPDs, arise due to the C-C between two adjacent pyrimidines and constitutes the majority of photoproducts formed (80-90%). The second major lesions are the pyrimidine – pyrimidone (6-4) photoproducts (Pyr [6-4] pyr), which are formed by the bonding of the C6 and C4 positions of two adjacent pyrimidines.

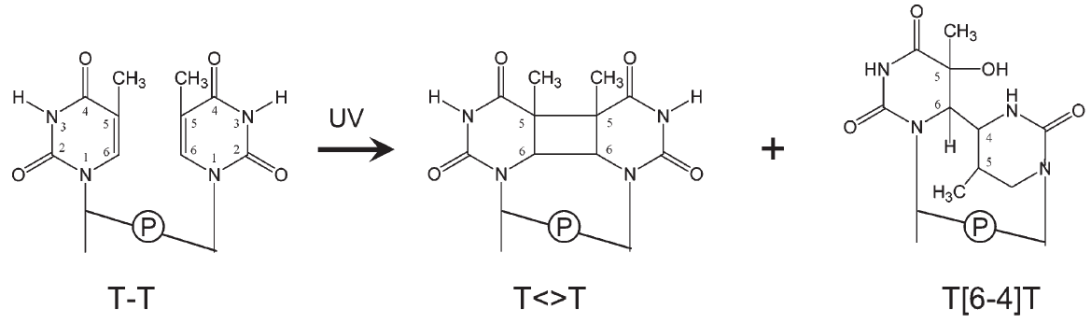


Figure 1.4: Structures of UV-induced DNA photoproducts. Cyclobutane pyrimidine dimers (T<>T) and pyrimidine (6-4) pyrimidone photoproducts (T[6-4]T). Reproduced from Sancar (2003).

Lesions in DNA induced by UV radiation are repaired by two broad pathways: light-dependent photoreactivation repair or light-independent dark repair (mainly through nucleotide excision repair). Among archaea, light-dependent photoreactivation has been demonstrated to be most efficient in repairing UV damage. In this process, a single, monomeric enzyme, a photolyase, directly photoreverses the UV-induced dimer through an electron transfer cycle using visible light as an energy source. This mechanism has been observed in both halophilic and hyperthermophilic species. For instance, a study on the rates of repair in *S. solfataricus* shows that 50% of the induced dimers were removed within the first 30 minutes with white light exposure, where the rate dropped 3 to 5 fold when the light was absent (Dorazi et al. 2007). Photoreactivation is especially well-characterized for *Halobacterium* sp. NRC-1, as studies have shown that repair in the light is far more efficient than in the dark (McCready et al. 2005). Figure 1.5 illustrates that in the presence of white light, *Halobacterium* can survive a much higher UV dose compared to the other key organisms, providing further evidence for the extreme efficiency of the repair

system. As the gene encoding the photoreactivatin enzyme is constitutively expressed, it is most likely the first line of defence against UV induced DNA damages.

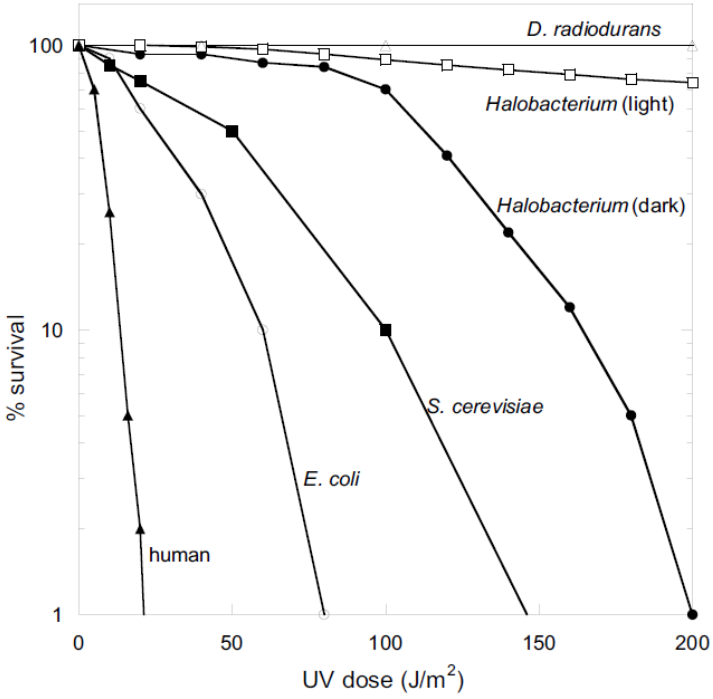


Figure 1.5: Survival curves for various organism with increasing levels of exposure to UV-C radiation. Percent survival is plotted vs. UV dose for human fibroblasts, *E. coli*, *S. cerevisiae*, *Halobacterium* in the dark or in the light, and *Deinococcus radiodurans*. Reproduced from McCready et al. (2005).

1.7 Nucleotide excision repair

In contrast to the light-dependent photoreactivation, excision repair pathways do not require light, and are classified into base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (Friedberg et al. 1995). For the purpose of this discussion, only NER will be described. NER was first identified in *E. coli* through the discovery that UV radiation resulted in the repair synthesis of short stretches of DNA, an indication that complete nucleotides were removed and not just the damaged base (Pettijohn and Hanawalt 1964; Setlow and Carrier 1964; Howard-Flanders et al. 1966). Subsequently, the proteins involved in this mechanism were purified and characterized through the isolation of UV-sensitive mutants (Sancar et al. 1981a; Sancar et al. 1981b; Sancar et al. 1981c). In the bacterial system, the uvrA, uvrB, and uvrC proteins constitute the uvrABC nuclease complex which is

required for the removal of pyrimidine dimers. The repair process begins by the probing of DNA abnormalities by uvrA, followed by the recruitment of the uvrB to the damaged region. The urvB forms a tight complex with the DNA at the site of the lesion and the uvrA dissociates. The uvrC now binds and starts the incision process 5' and 3' of the damaged site. Lastly, the helicase uvrD, removes the damaged oligos and the ligase Polymerase I restores the DNA strand (Sancar and Rupp 1983; Vanhouten 1990).

Humans and yeast provide two of the most well-characterized study models for the NER system in eukaryotes, where the Rad and XPS protein complexes were found to be involved in repair, respectively (Wood et al. 1988; Wang et al. 1993). Both systems utilize a host of proteins for the repair; Table 1.1 illustrates those known to date and their functions, demonstrating the complexicity of this mechanism in eukaryotes (Prakash and Prakash 2000).

Table 1.1 Yeast and human nucleotide excision proteins.

	S. cerevisiae gene	Human gene	Biochemical activities
	RAD7	Not known	Rad7–Rad16 complex, a DNA dependent
	RAD16	Not known	ATPase, binds UV-damaged DNA in an ATP dependent manner
	RAD14	*XPA*	Damage binding protein
	RAD4	*XPC*	Rad4–Rad23 complex binds UV damaged DNA
	RAD23	*HR23B*	
TFIIH	*RAD3*	*XPD*	5′ → 3′ DNA helicase
	RAD25	*XPB*	3′ → 5′ DNA helicase
	SSL1	*P44*	–
	TFB1	*P62*	–
	TFB2	*P52*	–
	TFB3	*MAT1*	–
	RAD1	*XPF*	Rad1–Rad10 nuclease cuts damaged DNA on the 5′-side of the lesion
	RAD10	*ERCC1*	
	RAD2	*XPG*	Rad2 nuclease cuts damaged DNA on the 3′-side of the lesion
	MMS19	Not known	None detected
	RAD26	*CSB*	DNA dependent ATPase
	RAD28	*CSA*	–

Reproduced from Prakash (2000).

Whilst gene homologues with the bacterial *uvrABC* are only present in mesophilic archaea (Grogan 2000), most archaeal species possess eukaryotic NER nucleases XPF (Rad1) / XPG (Rad2) and helicases XPB (Rad25) / XPD (Rad3). The occurrence of the uvrABC repair system in archaea is most likely a reflection of a lateral gene transfer event from the bacterial donor (Grogan 2000). However, little is known about the archaeal NER system. The damage recognition proteins XPA and XPC are absent from archaea, and for several species one or more of the eukaryal

NER enzymes are also missing. The published data on the full characterization of the NER pathway are limited to *M. thermautotrophicus* (Ogrunc et al. 1998) and *Halobacterium* sp. NRC-1 (Crowley et al. 2006), where full sets of the uvrABC proteins have been identified. In the halophilic archaea, deletion mutants lacking functional *uvr* genes are shown to be hypersensitive to UV and unable to remove CPD and (6-4) photoproducts post UV-C irradiation. Recently, two reports have confirmed a functional NER system in *S. solfataricus*, where the removal of photoproducts under permanently dark conditions is observed *in vivo* and *in vitro* (Dorazi et al. 2007; Romano et al. 2007). A complete pathway, however, has yet to be characterized.

1.8 Photoreactivation

1.8.1 Historical perspective

Photoreactivation was first discovered by Kelner in 1949, when he found that the lethal effects of UV irradiation on *Streptomyces griseus* were reversed when subsequently exposed to visible light (Kelner 1949). The same finding was obtained with *E. coli* where the author suggested that photoreactivation probably occurred in many organisms, including those of the higher forms (Kelner 1949). It was not until 1958 that Rupert and his colleagues finally discovered the enzyme responsible for such mechanism using a DNA transformation assay for *Haemophilus influenzae*. It was concluded that *E. coli* contained a light-activated enzyme that repaired UV-induced DNA damage, and was named photoreactivating enzyme (PRE), later known as photolyase (Rupert et al. 1958). Subsequently, Rupert established a reaction mechanism of PRE, showing that the reaction proceeded via the classical Michaelis-Menten kinetics and that visible light was absolutely essential for catalysis to occur. He also determined some of the fundamental enzymatic parameters of PRE using various *in vitro* studies (Harm and Rupert 1970a; Harm and Rupert 1970b). The identification of dimer formation in DNA was demonstrated shortly after the discovery of the photolyase, and that these dimers accounted for a large fraction of the effects of UV on DNA (Setlow and Setlow 1962; Setlow and Carrier 1964).

In the mid 1970s, Eker and colleagues attempted to purify photolyase from Baker's yeast and *E. coli*. While they were able to devise a purification scheme for

Streptomyces griseus, the purification quantity was constantly low and insufficient to carry out further experiments. Nevertheless, they were able to show that the absorption spectrum of the enzyme exhibited a marked absorption band in the region of 400 to 460 nm, an indication for the presence of a chromophore as reported earlier by Kelner (Kelner 1951; Eker and Fichtinger-Schepman 1975). The work with *S. griseus* revealed the association of a low molecular weight compound bound to the native protein, later identified as an 8-hydroxy-5-deazaflavin derivative (8-HDF) with an equimolar yield to the photolyase (Eker 1980; Eker et al. 1980; Eker et al. 1981). It was not until 1984 that Sancar and colleagues were able to purified large quantities of photolyase from *E. coli* and identified it as a flavoprotein containing two light-harvesting co-factors, or chromophores (Jorns et al. 1984; Sancar and Sancar 1984). By using a strain carrying the photolyase-overproducing plasmid, the group was able to develop an extensive purification procedure to purify several milligrams of homogeneous protein (Sancar et al. 1984a). Both the *E. coli* photolyase gene and protein sequences were also determined (Sancar et al. 1984b). Subsequently, a second chromophore was identified from *E. coli* and yeast as 5,10-methenyltetrahydrofolate (MTHF) (Johnson et al. 1988; Wang et al. 1988).

In the late 1990s, the crystal structures of both *E. coli* and *Anacystis nidulans* photolyase were finally characterized to atomic resolution, allowing for detailed analyses and comparison of the structures and the chromophores (Park et al. 1995; Tamada et al. 1997). The amino acid sequences of more than 50 photolyases are now known, from organisms which are distributed in all three kingdoms (Sancar 2003).

1.8.2 Mechanism of action

The reaction mechanism of photoreactivation and the role of the two chromophores have been studied extensively by Sancar and others using both *E. coli* and *A. nidulans* as models (Malhotra et al. 1992; Sancar 1994). The present commonly accepted model for the steps involved in the enzymatic reaction for CPD photolyases is shown in Figure 1.6. In simplistic terms, the enzyme binds to the dimerized DNA indepdendent of light, and flips the dimer out of the double helix into the active site cavity to make a stable ES complex. The folate molecule, either MTHF or 8-HDF,

absorbs a near-UV/blue-light photon, between 300-500nm, and transfers this excitation energy to the flavin in its active, reduced form ($FADH^-$). This $FADH^-$ donates one electron to the dimer, leading to the destabilization of the C5–C5′ and C6–C6′ bonds of the cyclobutane ring and the conversion of the thymine dimer into its orginal bases. The formation of the repaired thymine monomers is completed with the back transfer of an electron to the $FADH^\bullet$ radical and restoring the catalytic $FADH^-$ (Worthington et al. 2003). This hypothetical model was proposed more than 20 years ago (Sancar et al. 1987), and has since been supported through extensive biochemical data, spectroscopic analysis, computer modeling and structural determinations. Using femtosecond-resolved fluorescence technology, Sancar and colleagues observed the direct transfer of the electron from the excited flavin co-factor to the dimer, and the back transfer of the electron to the flavin, and concluded that the photocycle is completed in a subnanosecond scale with no net change in the redox state of the flavin chromophore (Kao et al. 2005). The mechanism of photoreaction has also attracted the attention of physicists and chemists, where numerous studies, such as the electrochemistry of the flavin cofactor and the quantum mechanics and molecular mechanics (QM/MM) of the splitting of the cyclobutane ring, have been published (DeRosa et al. 2005; Masson et al. 2009).

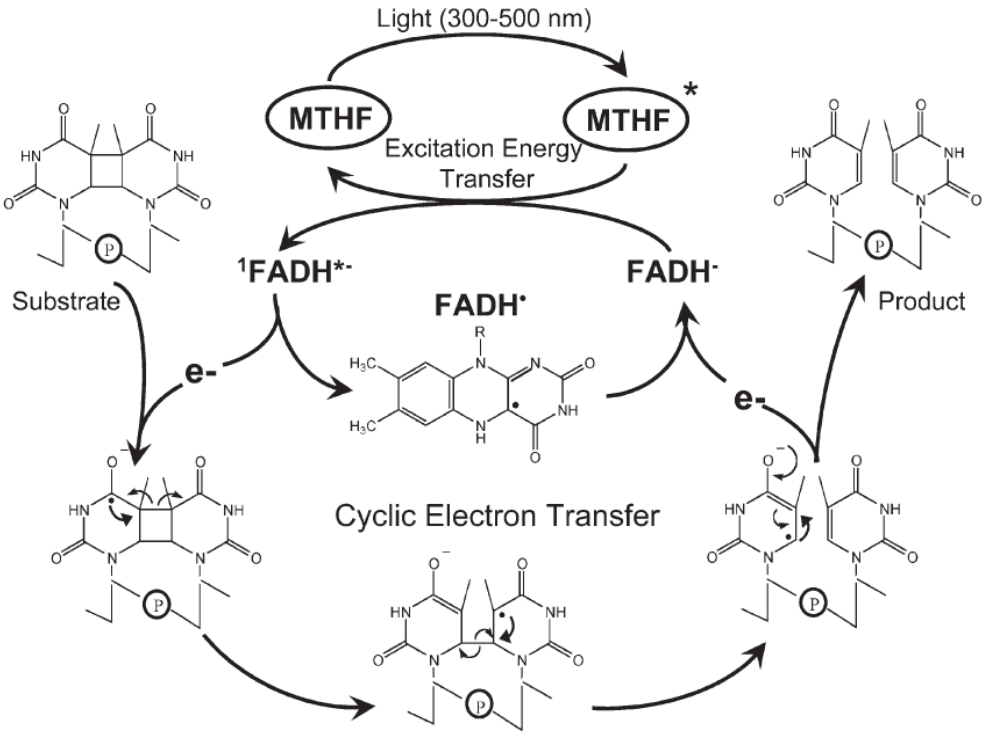


Figure 1.6: Reaction mechanism of DNA photolyase. The pathway is based on an *E. coli* photolyase. MTHF absorbs a 300-500nm nm photon, transfer the excitation energy to $FADH^-$, where the $(FADH^-)^*$ transfers an electron to Pyr<>Pyr and restores the intact DNA. Back electron transfer to the FADH° restores the catalytic FAD to its active form. Reproduced from Sancar (2003).

As the number of photolyase molecules per cell is relatively small, i.e. *E. coli* contains about 10-20 (Harm et al. 1968) molecules of CPD photolyase per cell, while yeast contains 75-100 (Yasui and Laskowski 1975), photolyases must be extremely efficient in distinguishing CPDs from a large excess of non-dimerized pyrimidine doublets. They are structure-specific DNA-binding proteins, meaning that the configuration of the DNA phosphodiester back bone at the damage site or the structure of the lesion itself determines specificity to the CPDs (Husain et al. 1987). A study using atomic force microscopy to visualize the conformation of DNA when *A. nidulans* CPD is bound to damaged DNA shows an average DNA binding angle of about 36° (van Noort et al. 1999). This led to the hypothesis that photolyase binds CPD by flipping out the lesion from the DNA double helix which is then stabilized by hydrophobic and electrostatic interactions (Vande Berg and Sancar 1998). This was demonstrated in *E. coli*, where fluorescent adenine analogues, 2-aminopurine (2-Ap), or 6MAP, were used to probe the local double helical structure upon binding of the substrate to the protein. Results indicated a dramatic change of the local

structure around the thymidine lesion when bound to photolyase (Christine et al. 2002; Yang et al. 2007). In addition, the resolved crystal structure of *A. nidulans* photolyase in complex with a model of the UV-damaged substrate showed that the dimer was flipped out of the DNA helix and into the enzyme active site (Mees et al. 2004). Figure 1.7 illustrates this base flipping phenomenon which has been observed for other enzymes as well, including T4 endonuclease V, which has been found to induce endonucleolytic cleavage of UV-irradiated DNA in a light-independent manner (McCullough et al. 1997; Goodsell 2005).

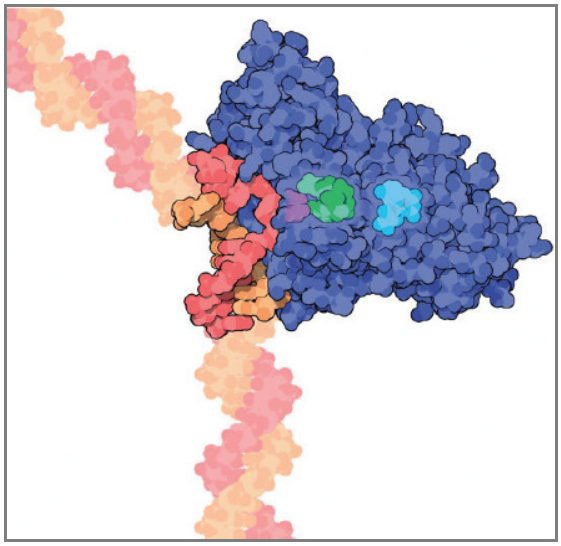


Figure 1.7: Base Flipping of CPD into the active site of photolyase. The structure of the DNA photolyase is shown in dark blue, DNA shown in red and orange, thymine dimer shown in magenta, FAD shown in green, and the photoantenna chromophore shown in light blue. Reproduced from Goodsell (2005).

The mechanism for (6-4) photolyase activity is not well understood due to the limited structural information on the enzyme; however, protein sequence homology to the CPD photolyase has provided a hypothetical model of the mechanism, shown in Figure 1.8. The mechanism is similar to that of the CPD photolyase, with a notable exception that upon binding to the substrate, the 3′ residue may become protonated by two histidine residues within the active site. This then facilitates a thermal reaction where the open form of the (6-4) photoproduct converts to the oxetane intermediate substrate. What follows is photoinduced electron transfer to the oxetane form of the (6-4) photoproduct to restore the two pyrimidines (Kim et al. 1994; Zhao et al. 1997). However, a recent report on the first crystal structure of a *Drosophila melanogaster* (6-4) photolyase in complex with a (6-4) lesion before

and after repair suggests a lack of the formation of the energetically unfavourable oxetane intermediate due to the position of and the inability of histidine residues for protonation (Maul et al. 2008).

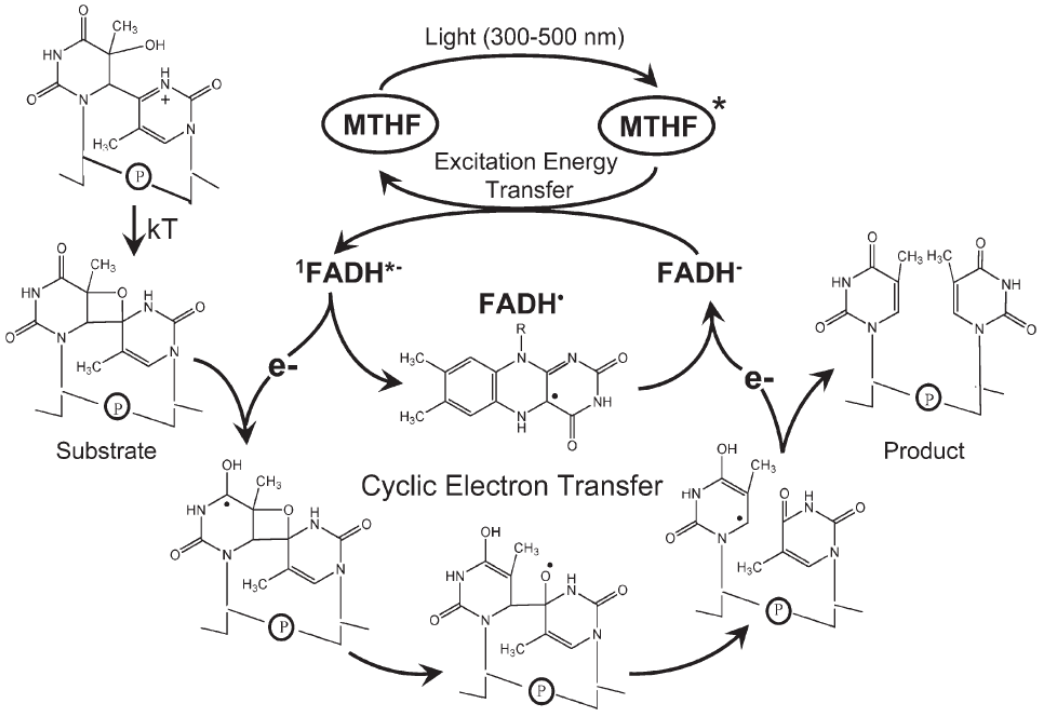


Figure 1.8: Reaction mechanism of (6-4) photolyase. Upon binding of the enzyme to the substrate, the 3′ residue becomes protonated via two histidines. This creates an oxetane intermediate that undergoes the same photochemical steps as the ones described for CPD photolyase. Reproduced from Sancar (2003).

1.9 Photolyase and cryptochrome superfamily

Traditionally, the photolyase/cryptochrome superfamily consists of five major subgroups based on sequence homology and known crystal structures. The five subgroups are: Class I CPD photolyases, class II CPD photolyases, plant cryptochromes, animal cryptochromes and 6-4 photolyases, and ssDNA photolyases (CRY-DASH). Recent advances in genomics and metagenomics have led to the increase in the number of photolyase/crytochrome family members, and now include 8 possible major subgroups including the new Class III photolyase, insect Cry1, and a bilateral crytochrome group consisting of insect Cry2 and vertebrate Cry (Ozturk et al. 2008). The most current phylogenetic analysis (Figure 1.9) revealed a previously undescribed prokaryotic group with unknown function, and 5 small divergent groups with narrow taxonomic ranges (Lucas-Lledo and Lynch 2009). This is a reflection

of the increase in the numbers of gene sequences as well as the growing sophistication of techniques used to re-construct evolutionary history. The following sections describe the five major subgroups that make up this photolyase/cryptochrome superfamily as well as the characterization of some new members.

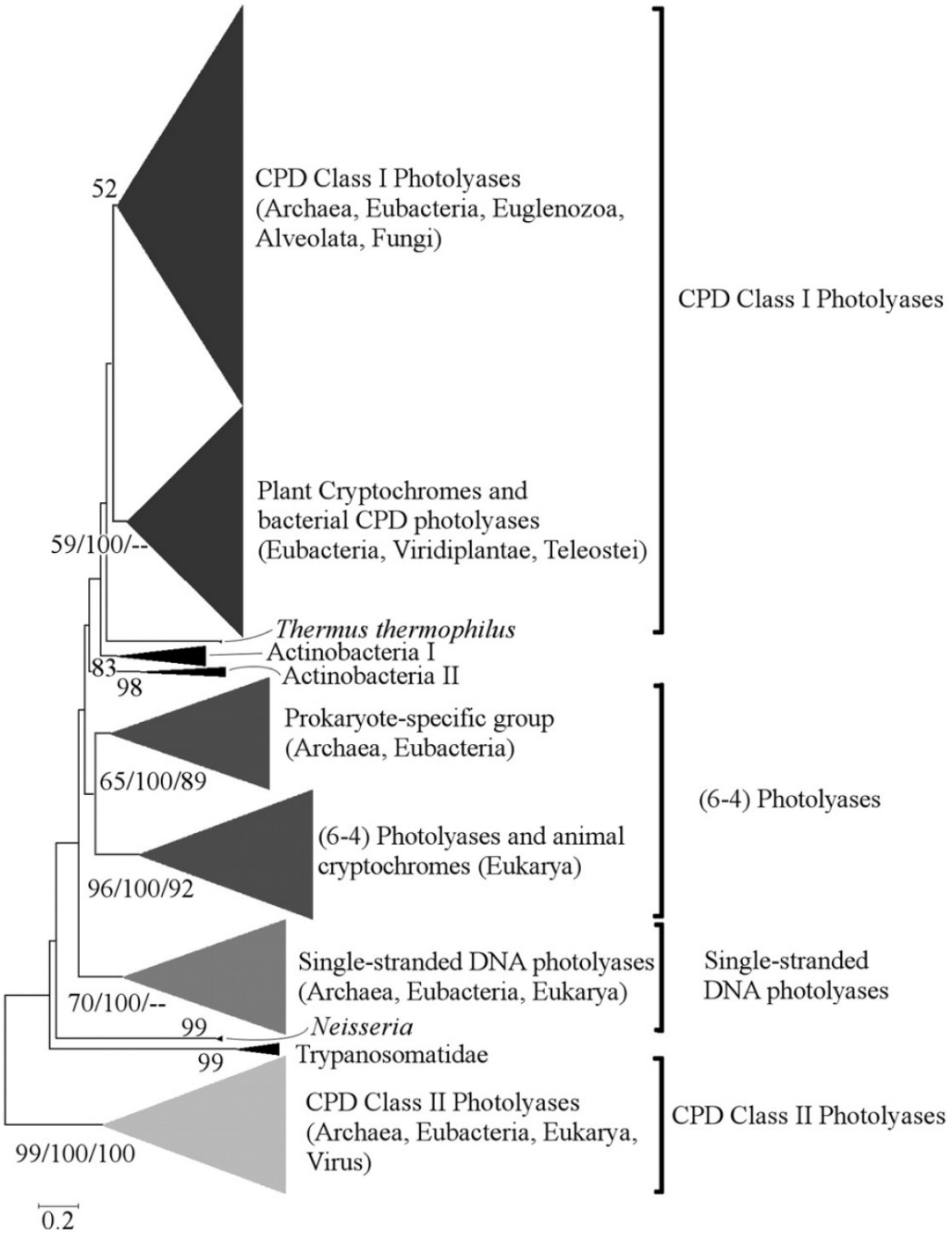


Figure 1.9: An unrooted phylogenetic tree of the photolyase/cryptochrome superfamily with subfamilies indicated on the right. Note that in this analysis the plant cryptochromes and the Class I CPD photolyases have been roughly grouped into one subclass, and that the (6-4) photolyase and the animal cryptochromes have been classified as the (6-4) photolyase subgroup. The single stranded-DNA photolyase refers to the CRY-DASH proteins. Adapted from Lucas-Lledo and Lynch (2009).

1.9.1 Photolyases

A photolyase is a monomeric enzyme of 450-550 amino acids, and consists of two noncovalently bound chromophores, or co-factors, first of which is FAD, while the second can either be MTHF or 8-HDF, depending on the organism. Based on the difference in the second light-harvesting chromophore, photolyases have been classified into either folate class (MTHF-type) or deazaflavin class (8-HDF-type) (Sancar and Sancar 1987). Figure 1.10 shows the near-UV/vis absorption spectra of the two chromophores. Photolyases of the deavaflavin class show the same absorption spectrum with a peak at 440nm, while those of the folate class display absorption maxima at a range between 370nm and 420nm. Diagrams of $FADH^-$ (active form of FAD) and the two light-harvesting chromophores are shown in Figure 1.11. Work by Sancar and colleagues has defined FAD as the essential chromophore for binding to the damaged DNA and for catalysis (Jorns et al. 1987; Sancar et al. 1987; Payne and Sancar 1990). The second chromophore, however, is not involved in catalysis or binding to the substrate; it has been shown that under limiting light, the second chromophore increases the rate of repair 10-100 fold (Jorns et al. 1984; Eker et al. 1988; Johnson et al. 1988). The apoenzyme lacking the 8-HDF chromophore has a very similar structure to that of the holoenzyme, indicating that 8-HDF is not essential for the folding of the protein either (Kort et al. 2004). Recently, FMN and FAD have been identified as the second chromophore for *Thermus thermophilius* and *Sulfolobus tokodaii*, respectively (Ueda et al. 2005; Fujihashi et al. 2007).

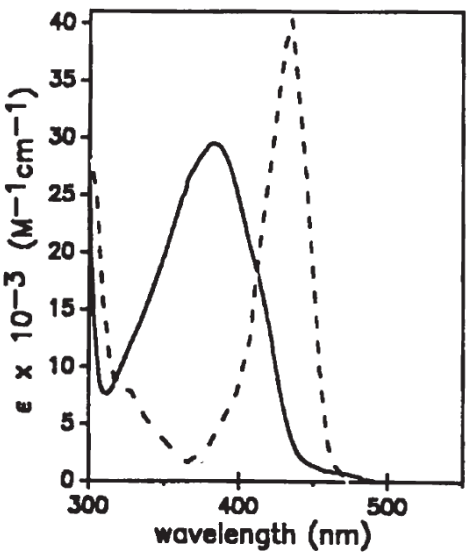


Figure 1.10: Absolute absorption spectra of folate and deazaflavin class photolyases in the near-UV/vis. Solid line is the absorption spectrum of *E. coli* photolyase (folate class). Dotted line is the absorption spectrum of *A. nidulans* photolyase (deavaflavin class). Reproduced from Sancar and Sancar (1987).

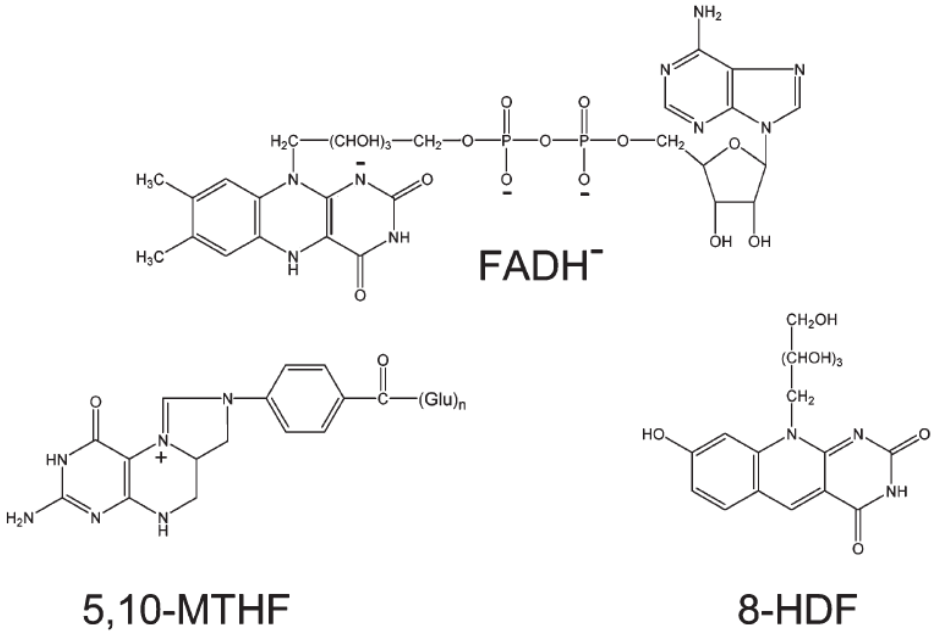


Figure 1.11: Structures of chromophores. FAD in its active two electron reduced form, $FADH^{\cdot}$ Reproduced from Sancar (2003).

Photolyases are distinguished by different substrate specificities; CPD photolyase binds to and repairs CPD lesions in single stranded or double stranded DNA, while (6-4) photolyase are specific for (6-4) photoproducts. CPD photolyases are further divided into two subclasses. In general, plant and animal CPD photolyases show a limited degree of homology to microbial photolyases in terms of amino acid

sequence; for this reason, they were divided into class I (microbial) and class II (animal) sequence groups. Class II photolyases have been isolated from higher organisms including the goldfish, *Carssius Auratus*, the fruit fly, *Drosophila melanogaster*; the rat kangaroo, *Potorous tridactylis*; the killerfish, *Oryzias latipes* (Yasuhira and Yasui 1992; Yasui et al. 1994); the archaea *Methanobacterium thermoautotrophicum* (Yasui et al. 1994); the eubacterium *Myxococcus Xanthus* (O'Connor et al. 1996); single celled algae *Chlamydomonas* (Petersen et al. 1999) and *Dunaliella salina* (Cheng et al. 2007); the plants *Arabidopsis thaliana* (Ahmad et al. 1997), *Cucumis sativus* (cucumber) (Takahashi et al. 2002) and *Oryza sativa* (rice) (Hirouchi et al. 2003) and the toad, *Xenopus laevis* (Tanida et al. 2005). This classification system, however, has drawbacks in that as more sequences become available, more diversity is observed amongst these sequences. For instance, a photolyase gene from the white mustard plant *Sinapis alba* is more closely related to the microbial class I than to the higher eukaryote class II photolyases (Batschauer 1993).

Of the two types of co-factors, MTHF has a maximum absorbance at 360nm which, upon binding to the enzyme, is red-shifted to 384nm. The folate class photolyases exhibit maximum catalytic activity at 377-410nm. Apart from *E. coli*, folate-type photolyases have been identified from *Salmonella typhimurium* (Li and Sancar 1991), *Vibrio cholera* (Worthington et al. 2003), the yeast *Saccharomyces cerevisiae* (Johnson et al. 1988), and the fungus *Neurospora crassa* (Eker et al. 1994). The deazaflavin class shows a maximum absorption range between 430nm and 460nm, where its absorption maximum of 438nm is also red-shifted to 420nm. Other than *A. nidulans*, the deazaflavin class of photolyases includes *Streptomyces griseus* (Eker et al. 1981), the green alga *Scenedesmus acutus* (Eker et al. 1988) and *Hbt. halobium* (Iwasa et al. 1988).

It was previously believed that the effectiveness of the excision repair system in removing pyrimidine (6-4) pyrimidone photoproducts in DNA, made the existence of a (6-4) photolyase redundant (Sancar and Rupp 1983). The CPD photolyases did not repair these types of DNA damage. However, (6-4) photolyases were isolated in *Drosophila*, followed by *Xenopus* and *Arabidopsis* (Todo et al. 1993; Kim et al. 1994; Kim et al. 1996; Nakajima et al. 1998), which catalyzed the photoreversal of

pyrimidine–pyrimidone (6-4) photoproducts. Sequence alignments of the *Drosophila* (6-4) photolyases and cryptochromes as well as a human homologue of (6-4) photolyase showed high sequence identity to the CPD photolyases, especially those in the Class I category which includes most of the microbial enzymes (Todo et al. 1996). This suggested structural similarity to the CPD photolyases. Subsequently, following the isolation of the *X. laevis* (6-4) photolyase, it was found to bind non-covalently to stoichiometric amount of FAD (Todo et al. 1997), leading to a proposed reaction mechanism for (6-4) photolyases similar to that of CPD photolyases (Zhao et al. 1997). Reaction mechanism of *X. laevis* (6-4) photolyase was investigated in detail using multiple mutant enzymes, showing the importance of two key histidine residues in the formation of the 4-membered ring intermediate in the repair process (Hitomi et al. 2001). More recent investigations on the crystal structures of the (6-4) photolyase from *D. melanogaster* and *A. thaliana* have allowed for a more comprehensive analysis on the degree of conservation between CPD and a (6-4) photolyases, as well as providing a better picture of the repair mechanism (Maul et al. 2008; Hitomi et al. 2009). Whilst the (6-4) photolyase has potential binding sites for both MTHF and 8-HDF, the binding site for the latter is better conserved, leading to the suggestion that the (6-4) photolyase more likely supports a flavin as its co-factor.

1.9.2 Cryptochromes

Cryptochromes (CRY) are proteins with high sequence homology to photolyases but have no apparent photorepair activity. Together with photolyases, they make up the photolyase/cryptochrome superfamily. Cryptochromes were first isolated from plants and were shown to mediate blue-light dependent growth and development (Ahmad and Cashmore 1993; Guo et al. 1998). Subsequently, cryptochromes were isolated from animals, where they regulate the circadian rhythm by light dependent and light independent mechanisms (Hsu et al. 1996; Kobayashi et al. 1998; Stanewsky et al. 1998; Thresher et al. 1998). These proteins, then, are loosely classified into the plant and animal subfamilies where the two are clearly distinguished in the phylogenetics tree. Cryptochromes exhibit 25-40% sequence identity to photolyases with a higher homology to (6-4) photolyases than to CPD photolyases (Thresher et al. 1998). It was later realized that (6-4) photolyases were

actually derived from the animal cryptochromes subfamily, clustering the two into one subfamily (Todo et al. 1996). While structurally cryptochromes and photolyases share a common conserved architecture containing both a FAD and a second folate chromophore, MTHF has not been detected in any of the resolved crystal structures of cryptochromes to date (Malhotra et al. 1995; Brudler et al. 2003; Brautigam et al. 2004).

A novel class within this superfamily emerged recently. Members of this class do not exhibit DNA repair activity either *in vivo* or *in vitro* and were previously classified as cryptochromes. Cry-DASH was first isolated from cyanobacterium *Synechocystis* (Hitomi et al. 2000) and *Vibrio cholera* (Worthington et al. 2003), followed by sequence homologues in *Arabidopsis thaliana* (Kleine et al. 2003), *X. laevis* and several other species. Based on sequence homologies, these were found to be more closely related to animal cryptochromes. To underscore the relationship of these cryptochromes from *Arabidopsis* and *Synechocystis* to the animal cryptochromes first isolated from *Drosophila* and Humans, this fifth subfamily was named CRY-DASH (Brudler et al. 2003). It came as a surprise when a later report indicated that these Cry-DASH proteins recognize CPDs within single stranded DNA and may contain residual photolyase activity (Daiyasu et al. 2004). This was clearly demonstrated with cryptochrome 1 from *V. cholera* (VcCry1) where it was found to be equally efficient in repairing damage in ssDNA as the photolyase (VcPhr2) in dsDNA. Sancar and colleagues attributed this phenomenon to lateral gene transfer, and suggested renaming these enzymes to ssDNA photolyases (Selby and Sancar 2006).

Members of the Class III subgroup constitute a sister class to plant cryptochromes and have been identified in more than 20 bacterial species, raising the possibility that these may constitute a new class of bacterial crytochrome whilst functioning as photolyases based on the exhibition of photolyase activity by at least four members of the group. The isolation and characterization of a Class III photolyase from *Caulobacter crescentus* have shown the enzyme is a bona fide photolyase that repairs CPDs and contains the MTHF chromophore (Ozturk et al. 2008). More recently, a PtCPF1 protein isolated from a marine diatom has demonstrated both (6-4) photolyase activity and role as a transcriptional repressor of the circadian clock in

a heterologous mammalian cell system. This could lead to a new subgroup within the superfamily, raising questions to the functional differentiation between photolyases and cryptochromes (Coesel et al. 2009).

1.9.3 Mammalian photolyases and cryptochromes

While sequence homologues of CPD photolyases have been identified in mammals, no photorepair activity has been observed. In 1974, a paper published in Nature described the purification and and characterization of a photolyase from human leukocytes (Sutherland 1974), and for more than two decades, the occurrence of photoreactivating activity in humans remained a vigorously contested subject. While photoreactivation was indeed demonstrated in marsupials (Cook and Regan 1969; Krishnan and Painter 1973), resulting in the cloning of the marsupial *phr*-genes and purification of a functional photolyase (Kato et al. 1994; Yasui et al. 1994), the original finding of a mammalian photolyase has yet to be replicated. It was demonstrated that microinjection of microbial photolyases from *A. nidulans* and *S. cerevisiae* into human fibroblasts followed by treatments with UV and visible light resulted in the successful removal of UV-induced CPDs within the human cells (Zwetsloot et al. 1985). After several failed attempts at cloning a human photolyase, it was finally determined that some time during evolution, placental mammals lost the photolyase gene, the reason for which is still unknown. It has been postulated that the loss of the photolyase gene in eukaryotic species may be due to weak natural selection, leading to a deleterious increase in genomic mutation rates (Lucas-Lledo and Lynch 2009). Finally, mammals, including humans, possess cryptochromes, a protein for which the structure and function have been well characterized (Ozturk et al. 2007; Tamanini et al. 2007).

1.10 Photoreactivation in halophilic archaea

In contrast to NER, photoreactivation is a well characterized mechanism in archaea, especially in halophilic species, where the encoding gene has been identified, and the gene product has been purified from a hyperthermophilic archaeon (Fujihashi et al. 2007). *Halobacterium* is a suitable model organism for the study of photoreactivation due to its extremely high tolerance to various stresses, including UV irradiation, in their natural habitats. Photoreactivation in halophilic archaea was first demonstrated in *Halobacterium cutirubrum*, where an exposure to visible light

after UV irradiation resulted in 100% revival of the strain (Hescox and Carlberg 1972; Fitt et al. 1983; Sharma et al. 1984; Eker et al. 1991). This was followed by findings that a dark repair pathway was absent in halophilic archaea (Grey and Fitt 1976). However, a study by McCready highlighted the efficiency of the light-dependent mechanism in the repair of UV-damaged DNA in *Halobacterium halobium* and *Haloferax volcanii*, as well as less efficient dark repair of both CPDs and (6-4) photolyases (McCready 1996). In an earlier study by Takao, a photolyase gene isolated from *Halobacterium cutirubrum* and expressed in *E. coli* was shown to increase the survival of UV irradiated host cells through photoreactivation (Takao et al. 1989). This was the first evidence that *phr2* encodes a CPD photolyase. According to the complete genome sequence of *Halobacterium* sp. NRC-1, two putative photolyase gene homologues, *phr1* and *phr2*, have been identified (Dassarma et al. 2001). McCready and Marcello reported the construction of a *phr2* deletion mutant, which was UV irradiated followed by subsequent exposure to visible light. The UV survival curve, as shown in Figure 1.12, clearly demonstrates the inability of the mutant strains to repair UV-induced damage, leading to a drastic decrease in its survival. The wild type strain exposed to visible light retained a high level of survival with increasing dose of UV up to 200 $J.m^{-2}$, while the unexposed wild type along with both the dark and light incubated mutant strains experienced comparable levels of decrease in survival. Through the use of a well characterized dot blot immunoassay, it was also determined that while (6-4) photoproducts were rapidly removed in the presence of white light for the *phr2* deletion strain, CPDs lesions remained. This gave strong evidence that *Halobacterium* possesses a (6-4) photolyase, and led to further confirmation that *phr2* encoded a functional CPD photolyase (McCready and Marcello 2003).

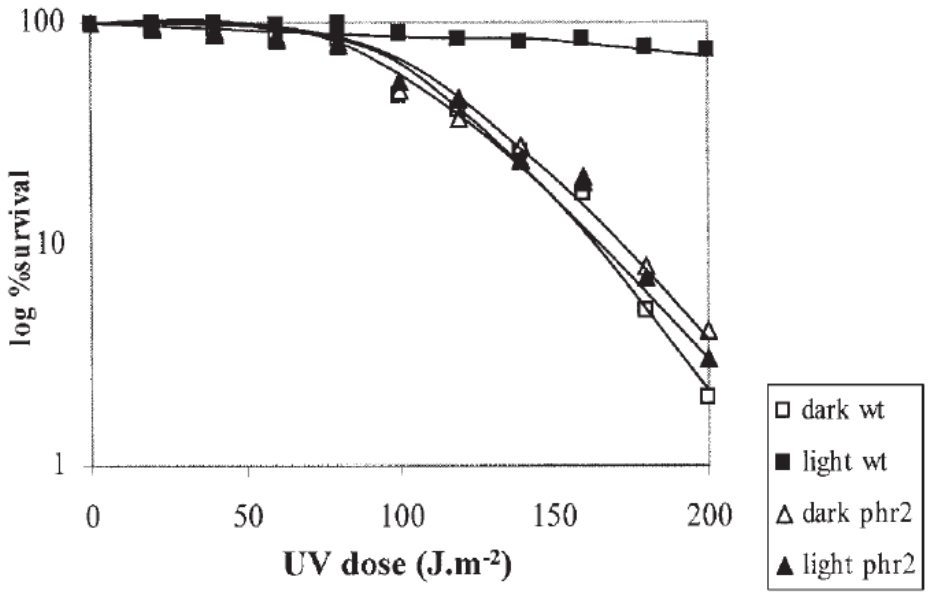


Figure 1.12: Survival curves of wild type and phr2 mutant strains with and without white light exposure post UV irradiation. Reproduced from McCready and Marcello (2003).

Baliga and colleagues utilized a systems approach to the elucidation of light and dark repair in *Halobacterium* sp. NRC-1 through the use of genetic analysis of both wild type and mutant strains in conjunction with a functional genomic approach. The group reached the same conclusion that *phr2* encodes a photolyse by looking at survival rates of both *phr1* and *phr2* mutants (Baliga et al. 2004). The analysis of mRNA changes during light and dark repair indicated the presence of additional putative repair mechanisms, as well as the observation of a UV-induced down-regulation of numerous essential metabolic functions. The other gene homologue, *phr1*, was not found to encode a photolyase and has been suggested to either function as a (6-4) photolyase or a light-harvesting cryptochrome. The function of *phr1* remains elusive today.

The crystal structure of the first archaeal photolyase was solved from *S. tokadaii*, revealing an overall conserved structure to other known photolyase structures. Photoreactivating activity of the purified photolyase was observed by measuring the absorption differences between normal thymine and dimerized thymine residues. The UV-irradiated DNA was mixed with the purified enzyme and incubated at room temperature under a fluorescent lamp. The ratio of recovered thymine concentration was calculated from absorption at 265nm at each time point. Figure 1.13 shows the

result obtained with the assay, indicating a higher percentage of repaired photoproducts for the photolyase-containg sample exposed to white light compared to the same sample incubated in dark only, and both light and dark treated non-photolyase-containing samples (Fujihashi et al. 2007).

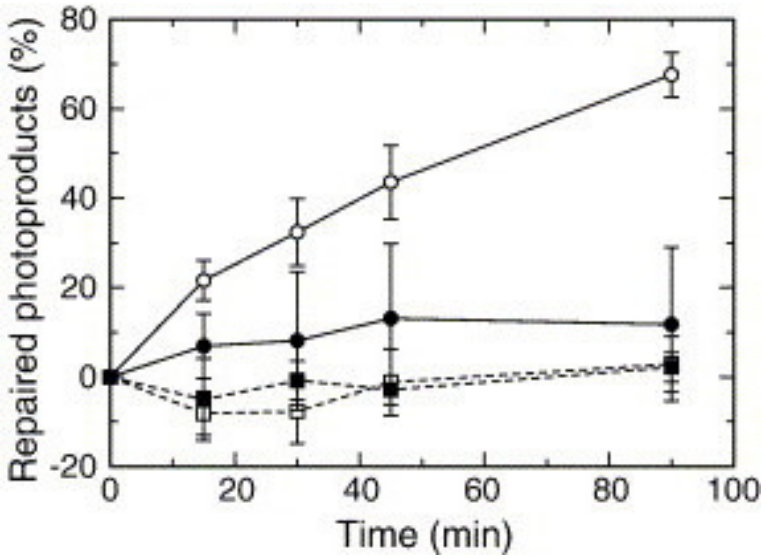


Figure 1.13: Photoreactivating activity of the purified photolyase from *S. tokodaii*. 4 different conditions were set up for this experiment, where the x-axis referred to incubation time either in light or dark. White light exposed sample containing photolyase is represented by open circle, while dark incubated photolyase-containing sample is represented by closed circles. Non-photolyase-containing mixtures in light and dark are indicated by open and closed square, respectively. Reproduced from Fujihashi et al. (2007).

More recently, it was found that the archaeal nucleohistone-like proteins, Sso7d and Sac7d are capable of repairing CPDs using an electron transfer process from the proteins to the DNA. This has implications in an evolutionary context, where nucleoproteins probably evolved to function in UV protection and DNA repair before the existence of the more efficient photolyase system, to allow survival in primitive Earth where the ozone layer was not present (Tashiro et al. 2006).

1.11 Project aim

While a few studies have indicated an enhancement of UV-survival due to white light illumination in *Hfx. volcanii*, the genomics and functional characterization of photoreactivation in this organism remains largely unexplored. The broad aim of this study was the identification of putative photolyase-like gene homologues, *phr1* and *phr2*, in *Hfx. volcanii* and the over-expression and functional characterization of the encoded proteins *in vivo* and *in vitro*. This study identifies a novel photolyase gene in *Hfx. volcanii* and provides the first demonstration that over-expression of the putative halophilic photolyase, Phr2, enhances tolerance to UV radiation. It also demonstrates that the light-harvesting chromophore required for photolyase activation in halophilic archaea is not MTHF but the deazaflavin, 8-HDF. To date, there have been no studies on the crystal structure of a halophilic photolyase. This study also explores the different routes to over-expressing and purifying a putative halophilic archaeal photolyase in homologous and heterologous expression systems, in an attempt to obtain enough soluble protein to enable crystal structure analysis.

These objectives are approached by:

- Identification of the putative gene homologues through bioinformatics analysis using the known gene sequences from *Halobacterium* sp. NRC-1
- Sequence alignment analyses and homology modeling of the *Hfx. volcanii* photolyase based on known structures of *E. coli* and *A. nidulans* photolyases
- Amplification, cloning and the homologous expression of the putative *phr* gene products in *Hfx. volcanii*
- Purification of the homologously expressed *phr* gene products
- Amplification, cloning and the heterlogous expression of the putative *phr* gene products in *E. coli*
- Purification and re-folding of the heterologously expressed *phr* gene products
- Assessment of photoreactivation both *in vivo* and *in vitro* using three assay systems: photorepair assay, cell viability assay, and a dot-blot immunoassay

Structure of a Putative Photolyase

2.1 Introduction

The original finding of photoreactivation in 1963 (Setlow et al. 1963), as well as the downstream elucidation of the photoreactivation mechanism of the *E. coli* photolyase, as conducted by Sancar and Jorns in the mid 90s (Jorns et al. 1990; Chanderkar and Jorns 1991; Sancar 1994), were significant findings which contributed immensely to the current understanding of how photolyases work. Since then, the biological and physical aspects of photolyases from various organisms in the three kingdoms of life have been characterized. Resolving the crystal structures of photolyases has been one of the most exciting areas of research as it provides important information on the chemistry and geometry of the two cofactors present, and allows for further analysis of the energy-transfer process between these two cofactors. The first photolyase to be resolved to atomic resolution was from *E. coli* (Park et al. 1995), followed by the structural elucidation from *Anacystis nidulans* (Tamada et al. 1997). In 2001, the X-ray crystal structure of a class I thermostable photolyase from *Thermus thermophilus* (Kato et al. 1997) was resolved (Komori et al. 2001). More recently, the first archaeal photolyase structure was determined from *Sulfolobus tokodaii* (Fujihashi et al. 2007). These photolyases all share a similar structure consisting of an N-terminal α/β domain and a C-terminal helical domain; the two domains are connected by a long interdomain loop that wraps around the α/β domain.

Sequence alignments of all known photolyases showed varying degrees of similarity, ranging from 15% to 70% or more sequence identity (Sancar 2000). Interestingly, the highest degree of similarity among all photolyases occurs at the C-terminal region of 150 amino acids, with around 30% sequence identity. This region was confirmed to be the FAD binding domain using both protein chemistry (Malhotra et al. 1992) and crystallography (Park et al. 1993). The light-harvesting co-factor, or chromophore, is usually at the N-terminal half, which is less conserved in general. X-ray crystallography of *E. coli* photolyase revealed that the light-harvesting cofactor is MTHF, bound to the exterior of the cleft between the helical and α/β domains and partially protrudes from the enzyme (Park et al. 1995). MTHF, or

methenyltetrahydrofolate, accounts for the light harvesting cofactor in the majority of photolyases analyzed to date. The *A. nidulans* photolyase, on the other hand, contains 8-hydroxy-5-deazariboflavin (8-HDF) as its second co-factor, and the enzyme has sufficient space to contain the co-factor at the interior cleft of the α/β domain, essentially burying the 8-HDF to the inside of the cleft (Tamada et al. 1997). The light-harvesting co-factor of *T. thermophilus* photolyase was unknown, as spectroscopic analysis indicated that the FAD was the only co-factor present when the photolyase was expressed in *E. coli*. However, the crystal structure suggested that a large cavity exists inside the cleft between the two domains, and had enough space to accommodate an 8-HDF, but not an MTHF at the corresponding position (Komori et al. 2001). In 2005, Ueda et al developed a crystallisation method capable of retaining the light-harvesting co-factor. Ultraviolet-visible absorption spectra, reverse-phase HPLC, and NMR analyses of the chromophores revealed that the second co-factor was flavin mononucleotide (FMN). It was also shown that FMN increased the light absorption efficiency of the purified protein (Ueda et al. 2005). Interestingly, the crystallography data from the archael *S. tokodaii* photolyase revealed that a second FAD is found at the position of the light-harvesting co-factor, although its general structure is similar to that of the *A. nidulans* photolyase (Fujihashi et al. 2007).

The aim of this chapter is to construct a model of the putative *Hfx. volcanii* photolyase based on sequence identity with the other known photolyases described above, specifically *E. coli* and *A. nidulans*, as these two photolyases contain different light-harvesting co-factors. The model will give important speculative information on the similarities between the predicted structure from *Hfx. volcanii* to the ones already known, as well as the type of light-harvesting co-factor it most likely contains. Blast searches of the putative *phr* genes from the unfinished *Hfx. volcanii* genome were performed, followed by sequence alignment analyses with known photolyase sequences to determine the degree of similarity between them. A 3-dimensional model of the putative *Hfx. volcanii* photolyase was presented, and the interactions of the surrounding residues with the light-harvesting co-factor were examined.

2.2 Searching for the putative *phr* genes

The sequencing of the *Hfx. volcanii* genome was incomplete in 2006 at the time this bioinformatics exercise was undertaken, complicating the identification of the putative *phr1* and *phr2* gene sequences. The genome sequencing project is now complete and publicly released on the UCSC archaeal genome browser; however, genome annotation is still incomplete. Identifying putative *phr1* and phr2 gene sequences in *Hfx. volcanii* began with the identification of putative *phr1* and *phr2* homologues, from the fully annotated genome of *Halobacterium* sp. NRC-1 (available at the University of Maryland website http://halo4.umbi.umd.edu/cgi-bin/haloweb/nrc1.pl). The DNA sequences were then translated using the SIXFRAME tool from the Biology Workbench website (http://worbench.sdsc.edu). Alternatively, the protein sequences can now be obtained from the GenBank website (http://www.ncbi.nlm.nih.gov/protein/15790367?report=genpept).

Putative *phr1* and phr2 gene sequences from *Halobacterium* were used to blast the unfinished genome of *Hfx. volcanii* (obtained from http://zdna2.umbi.umd.edu/~haloweb/hvo.html). These sequences were copied and pasted into the *Hfx. volcanii* browser window, and a BLAST search was initiated where all possible alignments were identified. The *Hfx. volcanii* sequences were compiled as contigs of various lengths; the known *phr* sequences were then aligned to the contigs at different positions. A scan through the alignments revealed that the best match for *phr1* occurred within contig 2975. The complete contig sequence was analysed, and a full length nucleic acid sequence for *phr1* was determined. The sequence alignments for *phr2* were slightly more complicated, as the most significant matches occurred in contig 2975, 3012 and 3124. A thorough analysis revealed that the *Hfx. volcanii phr2* was found in both contigs 3012 and 3124, with an overlapping region of 15 bases, as determined using the CLUSTAL W alignment tool from SDSC Biology Workbench (http://worbench.sdsc.edu). The complete *phr1* and *phr2* sequences were translated into 6 open reading frames using the SIXFRAME tool from Biology Workbench (Nucleic Acid Tool > SIXFRAME). All generated amino acid sequences were aligned against the known protein sequences from *Halobacterium* sp. NRC-1 to determine the putative Phr1 and Phr2 from *Hfx. volcanii*.

It is now much easier to obtain genome information on *Hfx. volcanii* through the UCSC archaeal genome browser (http://archaea.ucsc.edu/). From the individual genome webpage for *Hfx. volcanii*, a "BLAT" search can be conducted by pasting the known protein sequences from *Halobacterium* sp. NRC-1 into the window where a search query is generated, shown as either "browser" or "details." The browser link shows various relevant information based on the BLAT search generated. The newly annotated gene, HVO_2911, produced the most significant matches to *phr*2 from *Halobacterium* sp. NRC-1. The 3 links on the bottom of the page give the full protein and mRNA sequences as well as the genomic sequences near the gene. The predicted Phr2 sequence using this method matched that of the one generated using the previous method described. It was more difficult in obtaining the *Hfx. volcanii phr1* seqeunce using this browser, as HVO_2911 again came up as a match to the *phr1* from *Halobacterium* sp. NRC-1. To resolve this problem, gene sequences from other organisms were used to conduct the blast searches. This meant conducting a new search of proteins from the photolyase/crytochrome superfamily and identify those sequences that were most similar to the Phr2 from either *Hfx. volcanii* or *Halobacterium* sp. NRC-1. The cryptochrome 3 (CRY-3) sequence from *Arabidopsis thaliana* was used to produce the significant match to the putative *phr1* as indicated by HVO_2843, sequence which agreed with the one identified earlier. Both gene and protein sequences can be found in the Appendix (A1-A4).

2.3 Sequence alignment analyses

Sequence alignments of the putative Phr1 and Phr2 proteins from *Hfx. volcanii* against both known and putative photolyase and cryptochrome sequences from other organisms were performed to determine the degree of similarity amongst them. The degree of similarity was used as a determinant for constructing the hypothetical model of the *Hfx. volcanii* photolyase. Previous findings have indicated that the Phr1 protein from *Halobacterium* sp. NRC-1 did not encode a functional photolyase. It has also been postulated that Phr1 may function as a blue light photoreceptor, or cryptochrome, a sequence homologue to the photolyase but with no DNA repair activity. It would be of interest to see if the Phr1 from *Hfx. volcanii* shares more sequence similarity to cryptochromes than photolyases. Representative members from each of the photolyase/cryptochrome superfamily were used for this alignment;

all known sequences were obtained from the NCBI protein database except for the *Sulfolobus* species which were obtained from the Archaeal Genome Browser. The Phr1 and Phr2 sequences of *Halobacterium* sp. NRC-1 were obtained from the GenBank database, and those from *Hfx. volcanii* were determined as per Section 2.2.

The sequences were then inputted, as FASTA format, into the TCoffee program as provided by the Swiss Institute of Bioinformatics (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi). The completed alignments were then opened with GeneDoc, a multiple sequence alignment editor and analyzer program (http://www.nrbsc.org/gfx/genedoc/index.html). The program generated a colour coded alignment which can be found in the Appendix (A5); the different colours were representative of the degree of similarity amongst the aligned residues. The red, orange and yellow shading indicated identical, strongly similar, and weakly similar residues, respectively. From this alignment, the region of about 150 amino acids towards the C-terminal end showed the greatest degree of similarity. This was most likely the FAD binding domain, a well conserved region amongst all members of the cryptochrome/photolyase superfamily (Sancar 2003). Percentage identity for any pair of sequences was calculated and the table for this is presented in the Appendix (A6). In terms of sequence identity, *Hfx. volcanii* Phr2 is most closely related to *Halobacterium* sp. NRC-1 Phr2 (61%) and Phr1 (42%), followed by photolyases from *A. nidulans* (38%), *S. griseus* (35%), *E. coli* (33%), and *V. cholera* (31%). It also shares some sequence identity with the three *Sulfolobus* photolyases (28-29%), as well as the 6-4 photolyases (24-25%). The percent identity starts to drop off below 22 with comparison to the Class II CPD photolyases, crytochromes and CRY-DASH, where it shares the least sequence identity with photolyases from *D. melanogaster* (14%), *M. thermoautotrophicum* (13%), and *P. tridactylis* photolyase (12%). These results indicate that the *Hfx. volcanii* Phr2 is most likely a Class I CPD photolyase; however, at this point it cannot be certain whether it contains the 8-HDF chromophore as it also shares high sequence identity to the MTHF-type of Class I CPD photolyases.

The sequence identity results for *Hfx. volcanii* Phr1 are very different from that of Phr2, as it appears to share the most identity with the CRY-DASH proteins from *A. thaliana* (37%) and *V. cholerae* (32%), followed by *Halobacterium* sp. NRC-1 Phr1

(28%) and Phr2 (25%), and *S. griseus* photolyase (26%). It also shares some sequence identity between the range of 21 to 24% with the three *Sulfolobu*s photolyases, the Class I photolyases, and the 6-4 photolyases. *Hfx volcanii* Phr1 shares 11 to 19% sequence identity with members of the Class II CPD photolyase and all cryptochromes with the exception of the plant *S. alba*, with which it shares 23% identity. Whilst previous studies have suggested that Phr1 may encode a cryptochrome based on its high sequence homology to these proteins (Baliga et al. 2004), it is not the case shown here. The phylogenetic analysis strongly suggest the role of Phr1 in *Hfx. volcanii* as a CRY-DASH protein which has demonstrated photoreactivating capability on CPDs within single stranded DNA.

The dendrogram (Figure 2.1) shows the groupings of the photolyase/cryptochrome superfamily, constructed using Clustal X alignment and plotted in Treeview. The 5 subgroups (CPD Class I photolyases, plant cryptochromes, ssDNA photolyases, 6-4 photolyase and animal cryptochromes, and Class II CPD photolyases) can be seen as distinguished groups, with the exception of the *Sulfolobus* species which appear to form a subgroup of their own. *T. thermophilus* photolyase and *V. cholerae* cryptochrome also form a group on the bottom. Based on this tree, *Hfx. volcanii* Phr2 is clustered with the *Halobacterium* NRC-1 photolyases as well as the *A. nidulans* and *S. griseus* photolyases, suggesting that it may be a HDF-type Class I CPD photolyase. The *Hfx. volcanii* Phr1 is clustered with the two CRY-DASH proteins, suggesting its function as a ssDNA photolyase.

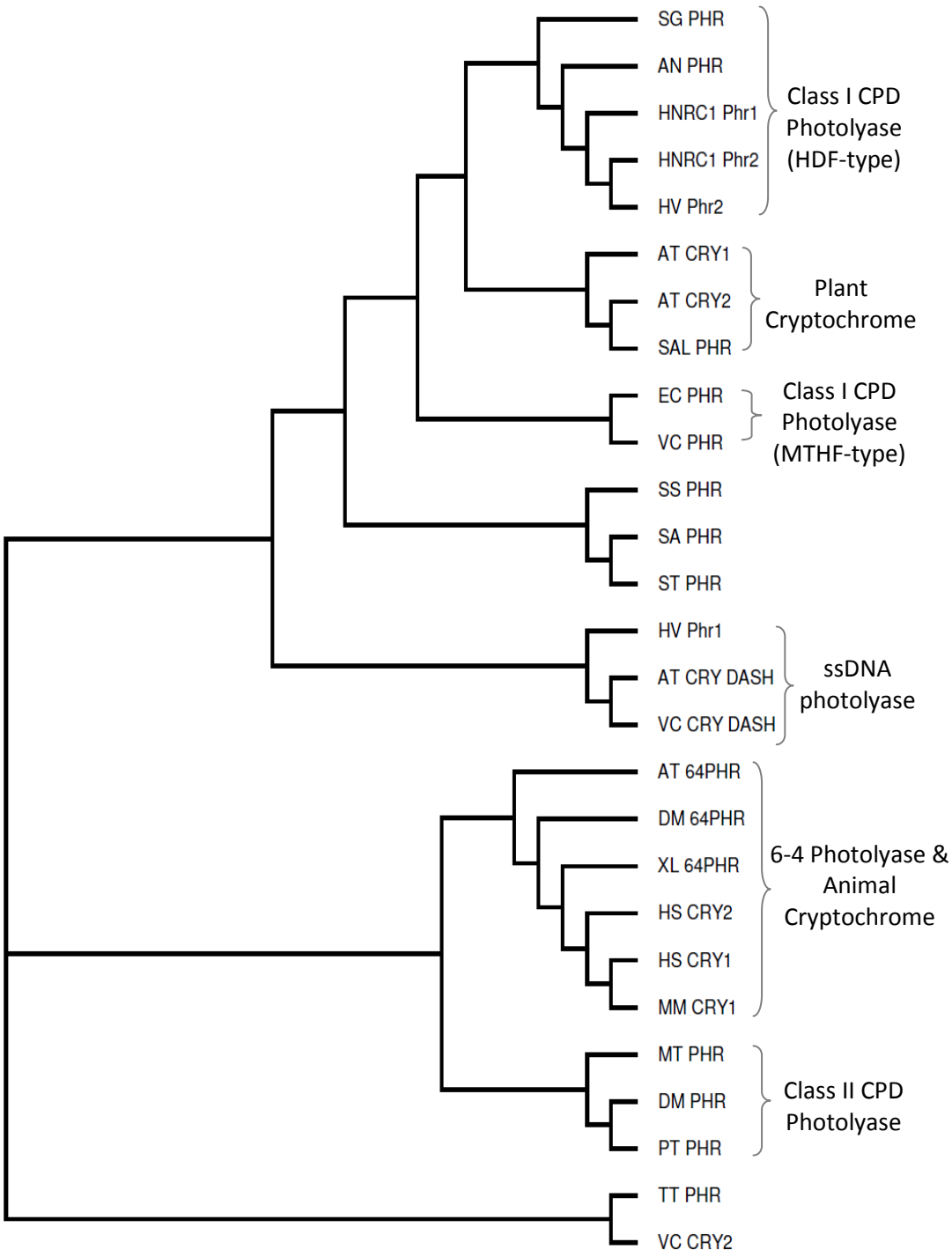


Figure 2.1: Non-bootstrap analysis of both putative and known photolyases
From top to bottom: *S. griseus* photolyase, *A. nidulans* photolyase, *Halobacterium* sp. NRC-1 Phr1 and Phr2, *Hfx. volcanii* Phr2, *A. thaliana* cryptochromes (CRY1 and CRY2), *S. alba* photolyase, *E. coli* photolyase, *V. cholerae* photolyase, *S. solfataricus* photolyase (putative), *S. acidocadarius* photolyase (putative), *S. tokadaii* photolyase, *Hfx. volcanii* Phr1, *A. thaliana* CRY-DASH, *V. cholerae* CRY-DASH, *A. thaliana* 6-4 photolyase, *D. manogaster* 6-4 photolyases, *X. laevis* 6-4 photolyase, *H. sapiens* crytochromes (hCRY1 and hCRY2), *M. musculus* cryptochrome (mouse), *M. thermoautotrophicum* photolyase, *D. melanogaster* photolyase, *P. tridactylis* photolyase, *T. thermophilus* photolyase and *V. cholerae* crytochrome.

A sequence alignment comparison of the 4 known crystal structures of photolyases and the *Hfx. volcanii* Phr2 is shown in Figure 2.2. Identical residues are highlighted in dark blue, while similar residues are highlighted in light blue. The residues interacting with the light harvesting co-factors 8-HDF, MTHF and FAD are coloured orange, purple and fluorescent green, respectively. The residues interacting with the catalytic FAD are coloured in pink. The yellow residues recognize CPDs, and the yellow residues with an orange highlight are the essential binding residues for dimers in *A. nidulans* photolyase. This alignment demonstrates that the FAD region is indeed well conserved amongst the 5 sequences compared, and it also shows that the *Hfx. volcanii* Phr2 has more in common with the *A. nidulans* photolyase in terms of possessing the same residues that interacted with the 8-HDF co-factor. These residues included arginine, valine, leucine, phenylalanine, aspartic acid and lysine. On the other hand, *Hfx. volcanii* Phr2 and *E. coli* photolyase did not share any of the same associating residues with MTHF. An initial deduction would be that the *Hfx. volcanii* Phr2 is more similar in structure to the *A. nidulans* photolyase, and that it probably contains 8-HDF as its light-harvesting co-factor.

```
                    *        20         *        40         *        60
H.volcanii  : MQ---LHWHRRDLRVADNRGLATAA--EAGPVAPLFVFDRDVLDHA--GAP-RVRYLLDALAELRG :  58
A.nidulans  : MAAPILFWHRRDLRLSDNIGLAAARA-QSAQLIGLFCLDPQILQSAD-MAPARVAYLQGCLQELQQ :  64
E.coli      : MTT-HLVWFRQDLRLHDNLALAAACRNSSARVLALYIATPRQWATHN-MSPRQAELINAQLNGLQI :  64
S.tokodaii  : MDC-I-FIFRRDLRLEDNTGLNYALS-ECDRVIPVFIADPRQLINNPYKSEFAVSFMINSLLELDD :  63
T.thermoph  : MGP-LLVWHRGDLRLHDHPALLEAL--ARGPVVGLVVLDPNNLKT----TPRRRAWFLENVRALRE :  59
              M    l w R DLR6 Dn  L  A       6  6   dp  l       p         6  L

                   *        80         *       100         *       120         *
H.volcanii  : AYQERGSDLLVARGDPRTVVPAVAAAFDAE----RAVWGIDYSGLARERDADVRLALDDAGVAREP : 120
A.nidulans  : RYQQAGSRLLLLQGDPQHLIPQLAQQLQAE----AVYWNQDIEPYGRDRDGQVAAALKTAGIRAVQ : 126
E.coli      : ALAEKGIPLLFREVDDFVASVEIVKQVCAENSVTHLFYNYQYEVNERARDVEVERALRN--VVCEG : 128
S.tokodaii  : ELRKKGSRLNVFFGEAEKVVSRFFNKVDA------IYVNEDYTPFSISRDEKIRKVCEENGIEFKA : 123
T.thermoph  : AYRARGGALWVLEGLPWEKVPEAARRLKA----KAVYALTSYTPYGRYRDAKVQEALPVP------ : 115
                   G  L    g             A            y    r RD  6  al

                   140         *       160         *       180         *       2
H.volcanii  : VHDAIFHPPGSITTNAGDTYSVYTYFWKKWRDRD-KPDPYP--EPDADDLVDAATLESAAEDLTNG : 183
A.nidulans  : LWDQLLHSPDQILSGSGNPYSVYGPFWKNWQAQP-KPTPVA--TPT--ELVDL-----SPEQLT-- : 180
E.coli      : FDDSVILPPGAVMTGNHEMYKVFTPFKNAWLKRLREGMPECVAAPK--VRS-SGSIEPSP-SITLN : 190
S.tokodaii  : YEDYLL-TPKSLF-----HHRNFTSFYNEVSKV-KVREPET--------------MEGSFD-VTDS : 167
T.thermoph  : --LHLLPAPHLLPPDLPRAYRVYTPFARRFLGVE-APLPAPEALPK--GPE--------------- : 161
                d 6   P  6       y v5t F           P     p                t

               00         *       220         *       240         *       260
H.volcanii  : DAEFDIAVGGLPTISDLGFEEP-SASVQPAGTEAARERLSAFCADAIYRYADDRDYPTRDATSRLS : 248
A.nidulans  : -AIAPLLLSELPTLKQLGFDWDGGFPV-EPGETAAIARLQEFCDRAIADYDPQRNFPAEAGTSGLS : 244
E.coli      : Y-------P------RQSFDTA-HFP---VEEKAAIAQLRQFCQNGAGEYEQQRDFPAVEGTSRLS : 239
S.tokodaii  : S---------MNVDFLLTFKK-IESPLFRGGRREGLYLLHRNV-----DF-RRRDYPAENNNYRLS : 217
T.thermoph  : -------------EGEIPRED-PGLPLPEPGEEAALAGLRAFLEAKLPAYAEERDRLDGEGGSRLS : 213
                          f      p    g  aa   L  f       5   R1 p     srLS

                   *       280         *       300         *       320         *
H.volcanii  : TDLKFGTIGIREVYAATAAARE-GVGGERDESVEEFQSQLAWREFYAHVLRGHPNVVT-ENYKEYE : 312
A.nidulans  : PALKFGAIGIRQAWRAASAAHALSRSDEARNSIRVWQQELAWREFYQHALYHFPSLAD-GPYRSLW : 309
E.coli      : ASLATGGLSPRQCLHRLLAEQPQALDG---GAGSVWLNELIWREFYRHLITYHPSLCKHRPFIAWT : 302
S.tokodaii  : PHLKFGTISMREAYYTQK-------------GKEEFVRELYWRDFFTLLAYYNPHVFG-HCYRREY : 269
T.thermoph  : PYFALGVLSPRLAAW--EAER---RGG---EGARKWVAELLWRDFSYHLLYHFPWMAE-RPLDPRF : 270
                l  G 6  R       a               5  2L WR F  h     P 6
```

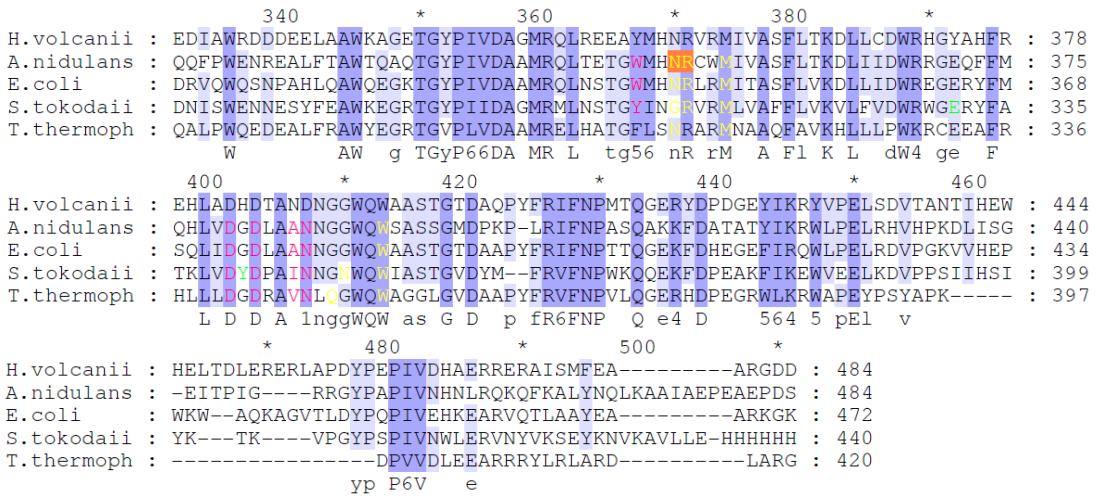


Figure 2.2: Sequence alignment analyses of the putative *Hfx. volcanii* Phr2 and known photolyases. From top to bottom: *Hfx. volcanii* Phr2, *A. nidulans* photolyase, *E. coli* photolyase, *S. tokodaii* photolyase and *T. thermophilus* photolyase. Identical and similar residues were highlighted in dark blue and light blue, respectively. The orange, purple and fluorescent green residues contacted the light harvesting co-factors 8-HDF, MTHF, and FAD, respectively. The catalytic FAD was co-ordinated by the pink residues. Yellow residues recognize the thymine dimers, and the yellow residues with an orange background indicate the essential binding residues for dimers in *A. nidulans*.

2.3 Structural modelling

A hypothetical model of the *Hfx. volcanii* Phr2 protein was constructed based on the known crystal structure of the *A. nidulans* photolyase as determined by the high sequence identity and the similarity of the binding residues between the two sequences. All diagrams were constructed using the PyMol program (http://pymol.sourceforge.net/). The modelling sequence in the .pdb format was provided by Dr. Susan Crennell, University of Bath. Figure 2.3 illustrates a complete model of the *Hfx. volcanii* Phr2 with the catalytic FAD co-factor as well as the two probable locations for the 8-HDF or the MTHF light harvesting co-factors. The co-factors are shown as the "sticks" structure in the protein. The overall structure of the photolyase consists of two domains: an α/β domain in the N-terminal region, and a helical domain in the C-terminal region consisting entirely of helixes and loops, both of which provide binding sites for either 8-HDF or MTHF, and FAD, respectively. This helical structure of the photolyase is similar to other members of the photolyase/cryptochrome superfamily. As seen from the model below, the FAD co-factor is buried within the centre of the helical domain, possibly making tight

associations with the photolyase. This was the case with both *E. coli* and *A. nidulans* photolyase, where the tightly-bound FAD could only be released after a mild denaturation of the protein (Jorns et al. 1990; Payne et al. 1990; Malhotra et al. 1992). The MTHF co-factor sits in a shallow groove between the α/β and the helical domain, whereas the 8-HDF is bound in a crevice formed in the α/β domain, shifted closer to the inside of the protein. The protruding location of MTHF as well as its ability to form hydrogen bonds with water molecules meant that it may dissociate more easily compared to FAD (Park et al. 1995) . This was observed for the *E. coli* photolyase, where after several column purification steps the protein contained substroichoimetric (20-30%) folate (Hamm-Alvarez et al. 1989; Hamm-Alvarez et al. 1990; Hamm-Alvarez et al. 1990).

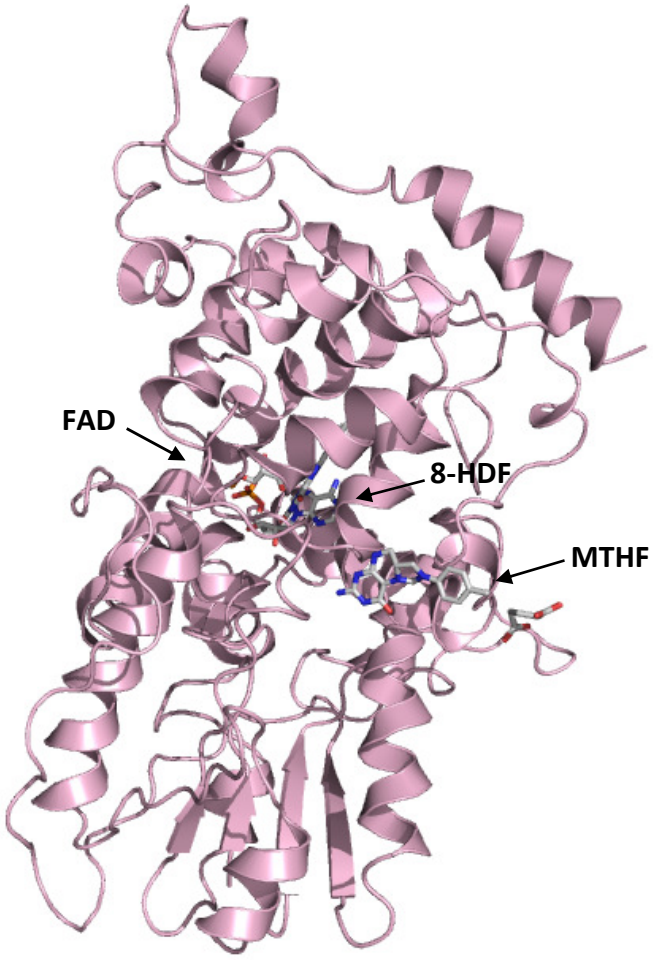


Figure 2.3: Structural Modelling of *Hfx. volcanii* with the catalytic FAD, and both 8-HDF and MTHF co-factors in their respective locations within the protein. The FAD co-factor was seen at the top of this model in the helical domain. The MTHF co-factor was shown to the right of the protein and partially protruding from it. The 8-HDF co-factor was situated within the α/β domain. Colour coding: Carbon = grey, Hydrogen = grey, Nitrogen = blue, Oxygen = red, and Sulphur = yellow

An analysis of the interacting side chains with either MTHF or 8-HDF was conducted to determine the type of light harvesting co-factor that the *Hfx. volcanii* photolyase most likely contained. Figure 2.4 illustrates superimposed images of photolyases from both *E. coli* (light orange) and *Hfx. volcanii* (light pink). As mentioned earlier, FAD binding residues are highly conserved amongst photolyases. It was only useful, then, to compare the MTHF binding residues from both structures. Figure 2.5 indicates the MTHF binding region from *E. coli*, where both the co-factor and the associated residues are shown and labelled. In *E. coli*, His44 forms a stacking interaction with the pterine ring of MTHF, while the side chains of Glu109 form hydrogen bonds with the nitrogens N1 and N3 of the pterdine ring, both of which are important interactions for binding the co-factor. Both the side chain oxygen and the peptide nitrogen of Asn108 are within hydrogen bonding distance with the hydroxyl group of the pterine. The side chain of Lys293 is especially important for forming a salt bridge with the Glu moiety of MTHF. A further contact occurs between the carbonyl oxygen of Cys292 and the positive charge on the five-member ring of MTHF, also found to be important for increasing affinity of the enzyme to the co-factor (Park et al. 1995).

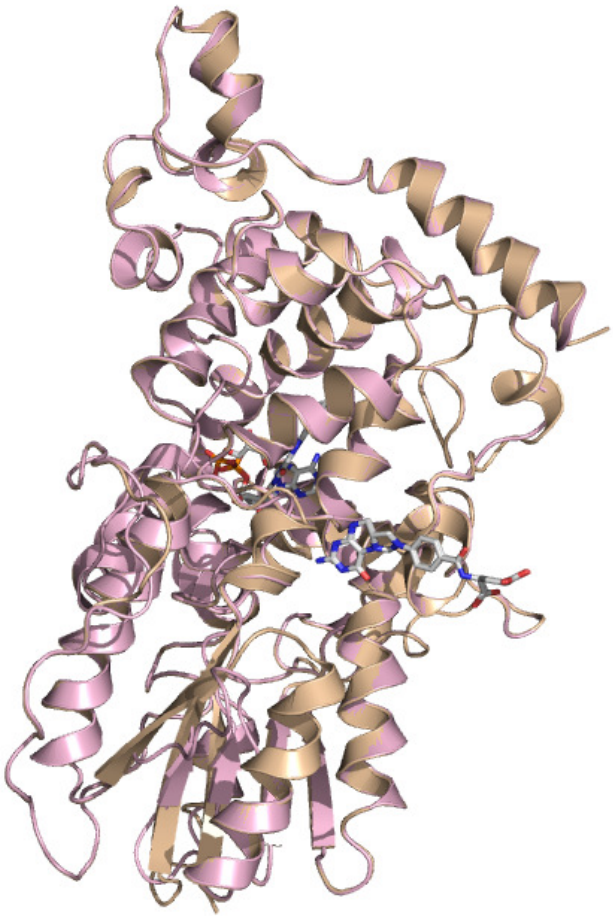


Figure 2.4: Superimposed images of *E. coli* and *Hfx. volcanii* photolyases. *E. coli* shown in light orange and *Hfx. volcanii* shown in light pink. Carbon = grey, Hydrogen = grey, Nitrogen = blue, Oxygen = red, and Sulphur = yellow

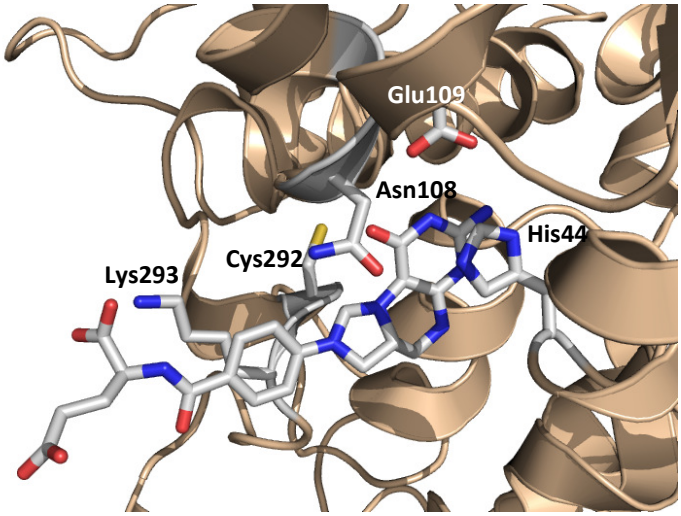


Figure 2.5: MTHF binding site of *E. coli* photolyase. A close-up view of the MTHF co-factor and the associating side chains. Carbon = grey, Hydrogen = grey, Nitrogen = blue, Oxygen = red, and sulphur = yellow.

Figure 2.6 below indicates the binding residues for *Hfx. volcanii* in the same positions as the ones described for *E. coli*. The residues His44, Asn 108, Glu109, Cys292 and Lys293 of *E. coli* were replaced with Ala41, Leu99, Ala100, Val304, and Thr305 in *Hfx. volcanii* in their respective positions. Major differences here occurred in the replacement of His44 by Ala41, a non-aromatic and non-polar residue. The negatively charged, polar Glu109 was replaced by a non-polar Ala100, while the polar Asn108 was replaced by a non-polar Leu99. Figure 2.6 shows a superimposed image of the residues from both photolyases, and it clearly demonstrates that the orientation of some of the residues is not correct to form interactions with MTHF; Thr305 is orientated away from the co-factor, meaning that it probably could not form a salt bridge like Lys293. The stacking interaction due to His44 is not present, and side chains of Ala100 would not form hydrogen bonds with the nitrogen atoms from the pterine ring. In addition, Leu99 cannot form hydrogen bonds with MTHF as its side chains do not contain either oxygen or nitrogen. Val304, in place of Cys292, forms the only possible van der Waals interaction with the MTHF co-factor in this figure. Judging by both sequence and structural alignments, it is unlikely that the *Hfx. volcanii* contains the folate type of light harvesting co-factor.

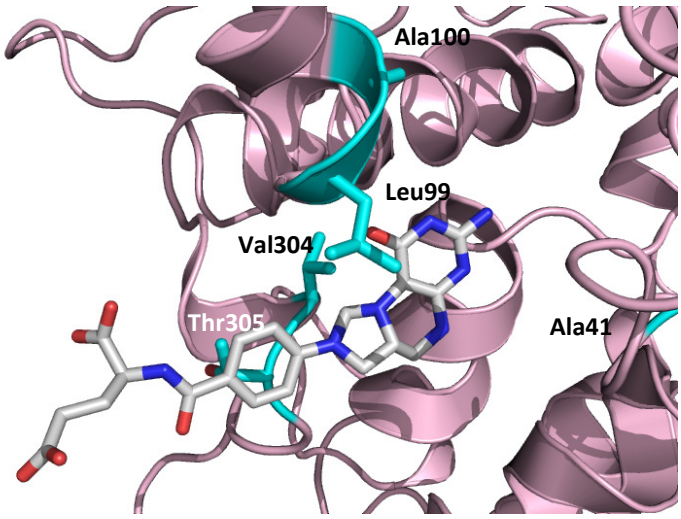


Figure 2.6: Potential MTHF binding site of *Hfx. volcanii* photolyase. A close-up view of the MTHF co-factor and the associating side chains. Carbon = cyan, Hydrogen = cyan, Nitrogen = blue, Oxygen = red, and Sulphur = yellow.

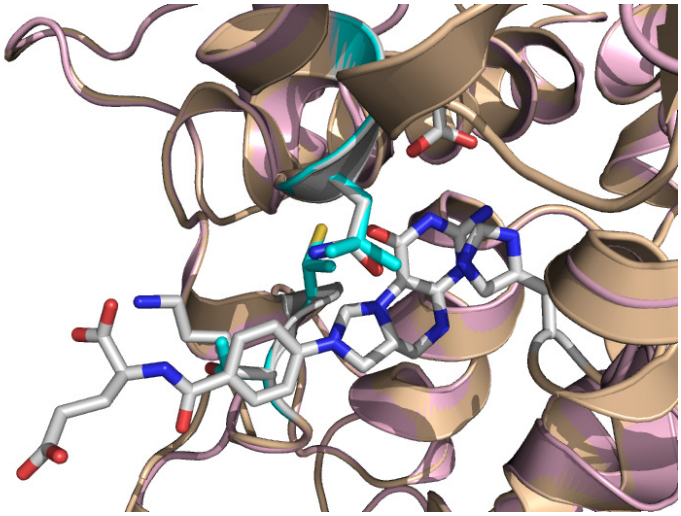


Figure 2.7: Superimposed images of the MTHF interacting residues from both *E. coli* and *Hfx. volcanii* photolyases. *E. coli* residues were shown in grey, and *Hfx. volcanii* residues were shown in cyan. Structure in the middle denoted MTHF co-factor.

A comparison of the side chains interacting with 8-HDF from *A. nidulans* (light yellow) and *Hfx. volcanii* (light pink) was conducted, where Figure 2.8 shows superimposed images of both structures including the FAD and the 8-HDF co-factors. Figure 2.9 illustrates a close-up view of the 8-HDF binding region and the residues found to be important for interaction (Tamada et al. 1997). The side chain oxygen atoms of Asp101 and Glu103 form hydrogen bonds to the O5' hydroxyl group of the 8-HDF ribityl moiety. The peptide nitrogen atom of Gly106 also forms the same interaction. The nitrogen atoms from the Arg109 side chain form salt bridges with the hydroxyl groups of the ribityl moiety of 8-HDF; it also forms a water-mediated hydrogen bond with the O_2 oxygen of the 8-HDF isozalloxazine ring. Cys36 and Asp38 form polar interactions with the oxygen and nitrogen atoms of the isoalloxazine ring. The side chain of Ile41 is at van der Waals contact distance to the isoalloxazine ring. Phe35 and Phe249 form non-covalent, stacking interactions while the side chain nitrogen atoms of Arg51 and Lys248 form hydrogen bonds with the 8-OH group of the isoalloxazine ring. Other contacting residues include Met47, Val52, and Leu55, all of which probably contribute to van der Waals interaction with the 8-HDF co-factor (Tamada et al. 1997). Residues interacting with 8-HDF are very well conserved in 8-HDF type of photolyases, where identical amino acids at positions Phe35, Asp38, Arg109 and Phe249 and

similar ones at Ile41, Arg51 and Lys248 are found in both *Hbt. halobium* and *S. griseus* (Yasui et al. 1994).

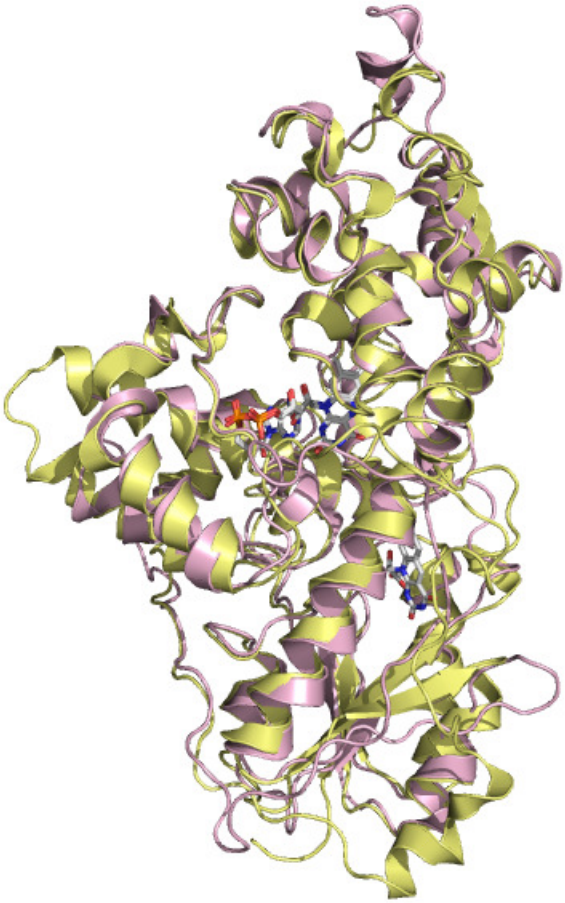


Figure 2.8: Superimposed images of *A. nidulans* and *Hfx. volcanii* photolyases. *A. nidulans* shown in light yellow and *Hfx. volcanii* in link pink. Carbon = grey, Hydrogen = grey, Nitrogen = blue, Oxygen = red, and Sulphur = yellow

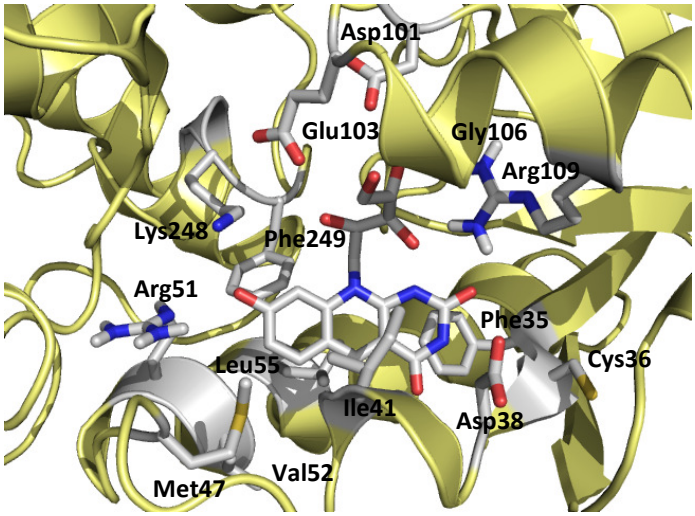


Figure 2.9: 8-HDF binding site of *A. nidulans* photolyase. A close-up view of the 8-HDF co-factor and the associating side chains. Carbon = grey, Hydrogen = grey, Nitrogen = blue, Oxygen = red, and sulphur = yellow.

Figure 2.10 shows the residues from *Hfx. volcanii* in the same positions as the ones described for *A. nidulans*, while Figure 2.11 is a superimposition of the residues from both species. The *Hfx. volcanii* Phr2 maintains the same aspartic acid residue (Asp95) in the same place, meaning that hydrogen bonding could still occur between the oxygen atom and the hydroxyl group; however, the replacement of Glu103 by Ser97 means that the side chain could no longer form a hydrogen bond with the hydroxyl group. The replacement of Gly106 by a structurally similar amino acid Ala100 means that the backbone amide nitrogen atom is still able to hydrogen bond with the hydroxyl group. The same salt bridge formation with the hydroxyl group and the water-mediated hydrogen bond with the isoalloxazine ring still occurs, as the *Hfx. volcanii* photolyase also contains an arginine (Arg103) in the same position as the *A. nidulans* photolyase. Although the replacement of Cys36 with a non-polar Val32 means that the polar interaction is absent, *Hfx. volcanii* photolyase maintains the same aspartic acid residue (Asp34) to form such interaction with both oxygen and nitrogen atoms of the isoalloxazine ring. Val37 forms van der Waals interaction with the isoalloxazine ring in the same way as Ile41. The positions of the phenylalanine residues are also conserved in *Hfx. volcanii* (Phe31 and Phe253), making the stacking interaction possible. Both Arg51 and Lys248 from *A. nidulans* are conserved as well in *Hfx. volcanii* (Arg45 and Lys252) to form hydrogen bonds between the side chain nitrogen atoms and the 8-OH group of the isoalloxazine ring. *Hfx. volcanii* photolyase contains Gly42, Val46 and Leu49, all of which are within

van der Waals contact distance to the isoalloxazine ring. Looking at the superimposed images of both photolyases, the majority of the residue side chains from *Hfx. volcanii* are oriented in the same way as those from *A. nidulans*, providing convincing evidence that the described interactions can occur. Based on the high identity and similaritiy of the 8-HDF interacting residues to the *A. nidulans* photolyase, the *Hfx. volcanii* photolyase probably contains 8-HDF as its second chromophore.

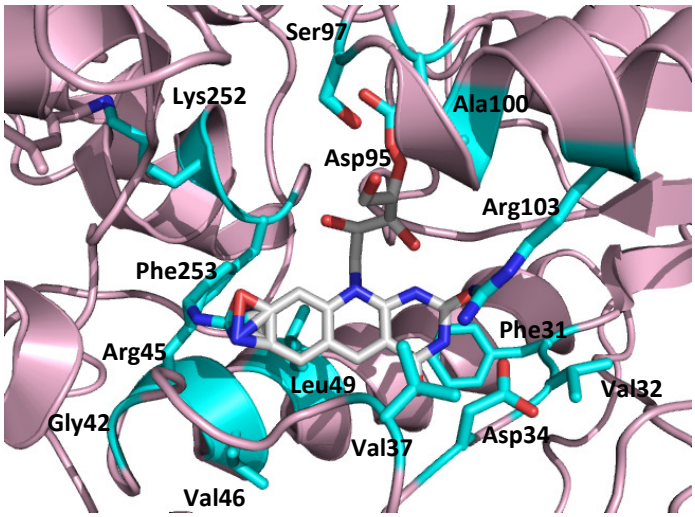


Figure 2.10: Potential binding site of *Hfx. volcanii* photolyase. A close-up view of the 8-HDF co-factor and the associating side chains. Carbon = cyan, Hydrogen = cyan, Nitrogen = blue, Oxygen = red, and Sulphur = yellow.

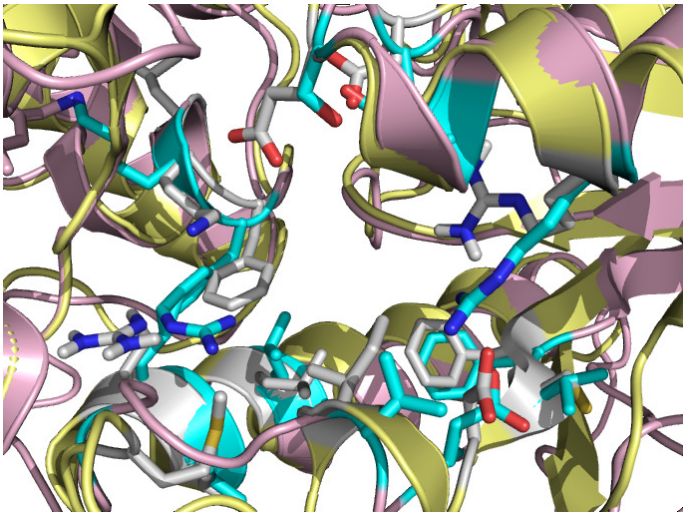


Figure 2.11: Superimposed images of the 8-HDF interacting residues from both *A. nidulans* and *Hfx. volcanii* photolyases. Grey residues were from *A. nidulans* and cyan residues were from *Hfx. volcanii*. The 8-HDF co-factor was not shown.

2.4 Discussion

Two putative photolyse-like gene homologues, *phr1* and *phr2*, were identified from the incomplete genome of *Hfx. volcanii* based on sequence identities to the known *phr* genes from *Halobacterium* sp. NRC-1. The dendrogram, consisting of representative members of the five major subgroups of the photolyase/cryptochrome superfamily, indicated that the *Hfx. volcanii* Phr2 falls under the class I CPD photolyases that are of the 8-HDF type, whereas Phr1 resembles that of the ssDNA photolyases. Judging by the sequence alignment analysis of *Hfx. volcanii* Phr2 with the known sequences of photolyase from *A. nidulans*, *E. coli*, *S. tokodaii* and *T. thermophilus*, it was also apparent the residues essential for binding FAD are present in the putative sequence of *Hfx. volcanii* Phr2. In addition, the residues important for the binding of 8-HDF appeared to be conserved between photolyases from *A. nidulans* and *Hfx. volcanii*, whilst the same conservation was not observed between *E. coli* and *Hfx. volcanii* for the binding of MTHF. As Phr2 is a more likely candidate as a photolyase than Phr1, a homology modelling of its structure, based on the known and resolved structures of *E. coli* and *A. nidulans*, was constructed. A close-up view of the residues and the orientation of these residues strongly suggest that *Hfx. volcanii* Phr2 binds 8-HDF and not MTHF. In the following chapters, this putative *phr2* sequence was amplified and cloned into appropriate vectors for the detection of the over-expression of the encoded protein either homologously or heterologously for the ultimate aim of purifying the protein. Functional analyses using various assays were also conducted for the elucidation of photoreactivating activity encoded by *phr2* from *Hfx. volcanii*.

Homologous Expression in Hfx. volcanii

3.1 Introduction

Recombinant protein expression in a heterologous strain, such as *E. coli*, has potential advantages over expression in the native organism. Rapid cell growth, high biomass levels and well developed molecular tools all contribute to high levels of expression of the desired protein. However, halophilic proteins require an environment of high ionic strength for their proper folding, activity and stability. The usual heterologous host strain *E. coli*, is a mesophile whose internal salt concentration is much lower than that of halophilic organisms. Previous research in the group has shown that the yield of *E. coli*-expressed halophilic enzymes, citrate synthase and DHlipDH, were greater than those obtained by purification from the native organism; however, these proteins were either soluble yet inactive, or insoluble inclusion bodies that required reactivation and refolding of the recombinant halophilic proteins (Connaris *et al*., 1999).

It is therefore desirable to express halophilic proteins in the native organism where the high salt concentration mediates proper folding and function of proteins (Eisenberg and Wachtel 1987). A polyethylene glycol (PEG) mediated transformation system was previously described and has since been optimized for use with various halophilic archaea species (Cline et al. 1995; Dyall-Smith 2008) A number of shuttle vectors have been developed which allow for replication in both *E. coli* and *Hfx. volcanii* due to the inclusion of the origins of replications for both species (Holmes and Dyall-Smith 1990; Holmes et al. 1994). There are also two types of selection markers for each host either for antibiotic resistance alone or antibiotic resistance and nutritional requirement (Allers et al. 2004).

Two gene homologs, *phr1* and *phr2*, have been identified from the fully annotated genome of *Halobacterium* sp NRC-1, where previous studies have looked at the phenotypic expression resulting from knocking out either or both of the genes (McCready and Marcello 2003; Baliga et al. 2004). The functional expression of the encoded proteins, however, has yet to be conducted. This chapter describes the extensive cloning procedure and the homologous expression of *phr 1* and *phr 2* from *Hfx. volcanii*.

3.2 Materials

3.2.1 Cell culture

Haloferax volcanii (DSM No. 3757) was obtained from DSMZ the Deutsche Sammlung von Mikrooganismen und Zellkulturen GmbH Braunschweig, Germany (German collection of Microorganisms and Cell Culture). Yeast extract, peptone and NaCl were supplied by Fisher Scientific, Loughborough, UK. Agar, ampicillin, carbenicillin, tryptone, polyethylene glycol (PEG, Average MW 600) and thymidine were from Sigma-Aldrich, Gillingham, UK. Bacto Yeast Extract, Bacto Casamino Acids and Bacto Agar were supplied by BD-Biosciences, Oxford, UK. Bacteriological peptone was obtained from Oxoid, Basingstoke, UK. Unless specified otherwise, all other reagents were ordered from Sigma-Aldrich, Gillingham, UK.

3.2.2 Molecular biology studies

Oligonucleotide primers were obtained from MWG-Biotech AG, Germany and Bioneer, USA. Taq polymerase, pGEM®-T Easy Vector system and JM109 competent cells were supplied by Promega, Southampton, UK. Restriction enzymes, Phusion™ High Fidelity DNA polymerase, $Vent_R$® DNA Polymerase, ThermolPol buffer, DMSO, and dam-/dcm- competent *E. coli* cells were obtained from New England Biolabs (NEB). 1 kb DNA ladder and high mass DNA ladder were obtained from Invitrogen, Paisley, UK. QIAEX II kit and the QIAprep Miniprep kit were provided by Qiagen, GmbH, Germany. Vectors pIL11, pTA233 and expression strain *Hfx. volcanii* H98 were provided by Dr. Thorsten Allers, University of Nottingham, UK. dNTPS, Isopropyl-β-D-thiogalactopyranoside (IPTG), and 5-bromo-4-chloro-3-indlyl-β-D-galactopyranoside (X-gal) were supplied by Sigma-Aldrich, Gillingham, UK. The protein standard was obtained from BioRad Laboratories, USA.

3.3 Methods

3.3.1 Strain condition, growth and preparation

3.3.1.1 *Hfx. volcanii* growth medium

Hfx. volcanii were grown in 18% (w/v) salt-water modified growth medium (MGM1) consisting of 14.4% (w/v) Nacl, 1.8% (w/v) $MgCl_2 \cdot 6H_2O$, 2.1% (w/v) $MgSO_4 \cdot 7H_2O$,

0.42% (w/v) KCl, 0.5% (w/v) peptone and 0.1% (w/v) yeast extract, pH 7.2, at 42 °C with shaking at 200 rpm . When required, the medium was solidified by the addition of 1.5% (w/v) agar (Dyall-Smith 2008). The plates were poured to a thickness of 5-8mm (in contrast to the normal 3mm), sealed in air tight containers to avoid dryness and salt crystal formation. The plates were placed in a non-shaking incubator at 42°C for 3 to 5 days for colony formation. Alternatively, the cultures may be left on the bench or in the refrigerator in well-sealed glass bottles or flasks (for liquid cultures) or on plates sealed in plastic. Using a sterile loop, the cells were inoculated into 50ml liquid medium in a 250ml conical flask and placed unto a shaking incubator at 37°C (at a recommended 1 to 5 ratio of medium to flask volume). Upon an OD_{650} reading of about 0.9, the cells were pelleted down by centrifuging in 50ml tubes at 4000rpm for 15 min and stored at -20°C for later use. Alternatively, the cells in liquid medium may be used for preparing glycerol stocks (3.3.1.3).

3.3.1.2 Genomic DNA extraction

Genomic DNA extraction for polymerase chain reaction was done by taking the cells directly from the plates using a disposable sterile loop and placed into 100µl sterilized water in a sterile 1.5ml microfuge tube. The above was done in the fumehood to avoid possible contamination of samples. The tubes were vortexed slightly, and then centrifuged in a tabletop microcentrifuge at 15,800 x g for 5 min (or 13,000 x g for 15 min). The supernatant containing the DNA was removed with a pipette and placed into a clean microfuge tube. The samples were stored on ice or into the -20°C freezer for later use. Alternatively, the cell pellets in liquid medium may be thawed, and using a sterile loop, re-suspended in 100µl sterilized water in a 0.5ml microfuge tube, then centrifuged at 13,000 x g for 15 min.

3.3.1.3 Glycerol stocks

800µl of cells (from liquid culture) were placed into small microfuge tubes and 200µl of autoclaved glycerol were added. The contents were mixed gently by flicking the tubes, followed by snap freezing the tube in an alcohol or dry ice bath. These glycerol stocks were stored in the -80°C freezer.

3.3.2 Cloning of *phr2* and *phr1* genes into pGEM®-T Easy Vector

3.3.2.1 Oligonucleotide primer design

Oligonucleotide primers were designed using the putative *phr2* sequence identified from *Hfx. volcanii*. The primers shared homology upstream and downstream of the *phr2* gene sequence. Both the forward and reverse primer incorporated restriction sites, *Nco*I and *Hin*dIII, respectively. These restriction sites were selected according to the pIL11 expression vector map sequence. To avoid any internal cut sites within the gene, i.e., to verify that the *Nco*I and *Hin*dIII sequences do not exist within the *phr2* gene, the restriction sites were mapped using a mapping program (www.restrictionmapper.com). The primers were checked for their melting temperatures, GC content, and the existence of hairpin or dimer structures using the Primerselect Programme (Primer Premier, version 4.04, Premier Biosoft International, Palo Alto, CA, USA). The forward primer had to have 2 base changes in order to incorporate the *Nco*I restriction site. The reverse primer was designed to include the *Hin*dIII restriction site. A reverse primer was also designed to check for the incorporation of the ferredoxin promoter in the *phr2*-pTA233 construct. The primer design for *phr1* was done accordingly, where restriction sites *Xba*I and *Hin*dIII were used for cloning purposes. There were no base pair changes as *Xba*I did not contain ′ATG′ in its sequence. The *phr* gene sequences, the flanking sequences to the genes as well as the primer sequences can be found in A1 and A2.

3.3.2.2 Polymerase chain reaction

The amount of the template *Hfx. volcanii* genomic DNA and the primers were 100ng and 100pmol, respectively. PCR reactions were set up in either 0.2 or 0.5 ml thin walled tubes on ice, and the reagents from Table 3.1 were added sequentially starting with water. The procedure was performed in the fumehood to minimize contamination. A control PCR tube that did not contain template DNA was used to detect the sterility of the reaction. The tubes were placed directly in a pre-heated Mastercycler thermal cycler (Eppendorf), where temperature cycling was carried out under the conditions described in Table 3.2. As the extension time is usually determined by the size of the expected product (1 min per kilobase); with *Hfx. volcanii* 3 min was always used. The annealing temperature was always 5°C below

the Tm of the forward and reverse primers; in this case they were 74.9°C and 75.3°C, respectively.

Table 3.1: PCR reagents

Reagents	Negative control	Single primer (Forward)	Single primer (Reverse)	Double primer
Autoclaved water	37	37	37	36
dNTPs	1	1	1	1
10X thermol buffer	5	5	5	5
DMSO	4	4	4	4
Forward primer	1	1	–	1
Reverse primer	1	–	1	1
Template DNA	–	1	1	1
Vent polymerase	1	1	1	1
Final volume (µl)	50	50	50	50

Table 3.2: PCR thermocycling conditions

PCR cycle conditions	Temperature	Time	
Step 1	96°C	5 min	Initial denaturation
Step 2	96°C	1 min 30 s	Denaturation
Step 3	70°C	1 min 15 s	Annealing
Step 4	72°C	3 min	Extension
***Go to step 2, repeat 30 cycles**			
Step 5	72°C	10 min	Final extension
Step 6	4°C		Hold

3.3.2.3 Agarose gel extraction

The PCR products were visualized on a 1% (w/v) agarose gel using 1% TAE buffer. Depending on the comb sizes, 10 or 20µl of samples mixed with a 6X dye was loaded into the wells. The gel was run at 70V for 1.5 hr, taken out of gel cast and visualized on an UV transilluminator. Using a sterile razor, the band containing the gene was excised and placed into a 1.5ml microcentrifuge tube. The extraction of DNA was performed using a QIAquick Gel Extraction Kit (Qiagen). 5µl of the eluted DNA was run on a 1% (w/v) agarose gel along with a high mass ladder to estimate the amount of DNA.

3.3.2.4 A-tailing of PCR product

Thermostable $Vent_R^®$ DNA polymerase generated blunt-ended fragments during PCR amplification; therefore, the PCR fragment needed to be modified by the addition of adenines to both ends before cloning into the pGEM®-T vector. The reaction was performed in 0.5ml thin walled tube by the addition of 5.6µl of the extraction product, 1µl 10 x reaction buffer free of $MgCl_2$, 2mM of $MgCl_2$, 2mM of dATP and 5 units of Taq polymerase to a final volume of 10µl. The reactions were incubated at 70°C for 30 min.

3.3.2.5 Ligation into pGEM®-T Easy Vector

The pGEM®-T Easy Vector System (Figure 3.1) was used to for the cloning of the A-tailed products, following the protocol as outlined in the Promega technical manual of pGEM®-T and pGEM®-T Easy vector Systems. The appropriate amount of eluted insert DNA was ligated into 1µl pGEM®-T Easy Vector (50ng) in 10-20µl of final volume made up with autoclaved MilliQ water (Table 3.3). The concentration of extracted DNA was estimated by comparison with the high mass ladder on the gel. The amount of extracted DNA to be used in the ligation reaction was calculated as below:

$$\frac{\text{ng of vector (50)} \times \text{kb size of insert (1.455)}}{\text{kb size of vector (3)}} \times \text{insert: vector molar ratio*} = \text{ng of insert}$$

*A 3:1 ratio of insert to vector was used when calculating the amount of insert.

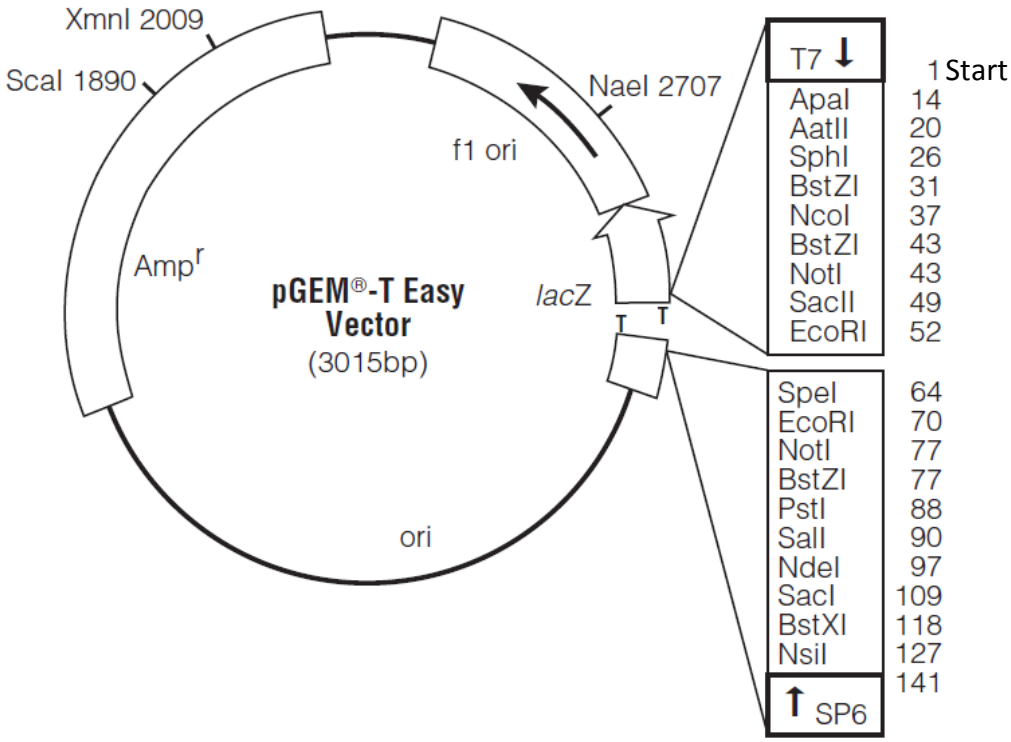


Figure 3.1: Circular map of pGEM®-T Easy Vector
Amp^r denotes the ampicillin/carbenicillin resistance gene. The T-overhangs are indicated in the *Lac Z* region, which is disrupted upon the insertion of a DNA fragment and forms the basis of blue/white screening. f1 ori is the origin of replication in *E.coli*. Inserts can be sequenced using SP6 and T7 promoter primers. Reproduced from the Promega pGEM®-T and pGEM®-T Easy Vector System Manual.

Table 3.3: Ligation reaction set-up

Reagents	Reaction 1	Reaction 2 (control)
2X rapid ligation buffer T4 ligase	5	5
pGEM-T Easy Vector (50ng)	1	1
DNA	(determined amount)	–
Control insert DNA	–	1
T4 DNA ligase (3 Weiss unit/µl)	1	1
Autoclaved MilliQ water	(add to final volume)	2
Final volume (µl)	10-20	10

3.3.2.6 Transformation of competent JM 109 *E. coli* cells

Following the protocol as outlined in the Promega technical manual, 10µl of the ligation reaction was used to transform 50µl of competent JM 109 cells. The LB plates containing either 50 µg/ml of carbenicillin or 100 µg/ml of ampicillin were made according to the protocol (unused plates were stored in the cold room for up to one month). The SOC medium was also made following the protocol manual, with the exception of adjusting the pH before autoclaving to avoid contamination. The selection plates were warmed prior to plating to allow easier spreading. Each plate was spread with 100µl of 100mM IPTG and 20µl of 50mg/ml of X-Gal and left to dry. 100µl of transformants were spread unto a plate labeled "dilute." The rest of the cells were centrifuged briefly, and 700µl of the supernatant was removed. The pellet was re-suspended again with the remaining supernatant, and 100µl was spread unto a plate labeled "concentrate." The plates were incubated at 37°C overnight to allow for blue/white selection. The white colonies were selected and streaked unto a master plate. Two or three blue colonies were also selected as controls (non-transformants).

3.3.2.7 Purification of plasmid DNA

5ml of LB medium containing 5µl of 100µg/ml ampicillin (or 5µl of 50µg/ml carbenicillin) was inoculated with a single colony, followed by incubation overnight in a 37°C shaking incubator at 150 rpm. It is not recommended to extend the incubation period for longer than 16 hr (QIAprep Spin Miniprep Kit manual). The cells were spun down at 6000 rpm for 8 min at room temperature. The QIAprep Spin Miniprep Kit was used to purify the plasmid DNA using the microcentrifuge method. The eluted plasmid DNA was visualized on a 1% (w/v) agarose gel for analysis.

3.3.2.8 Sequencing

Successful transformants were sent for sequencing using the SP6 and T7 primers. The samples were sequenced by GeneServices, Cambridge. The sequencing results were analyzed using the Chromas Lite program as well as the BioEdit Sequence Alignment Editor Software package.

3.3.2.9 Single and double digests

Single and double digests were performed to determine if the insert of the correct size is present. Components from Table 2.5 were added to a 1.5ml tube and incubated in a 37°C water bath for 4 hr. The reactions were visualized on a 1% (w/v) agarose gel alongside a DNA mass ladder to determine the sizes of the bands.

Table 3.4: Reaction set up for single and double digests

Reagents	Single digest	Single digest	Double digest
Plasmid DNA	5	5	5
Buffer II	—	5	5
Buffer IV	5	—	—
10X BSA	5	5	5
Enzyme 1	1	—	1.2
Enzyme 2	—	1	1.2
Sterile water	35	35	32.6
Total volume (µl)	50	50	50

3.3.3 Cloning of *phr* genes into pIL11 and pTA233 vectors

3.3.3.1 pIL11 and pTA233 vector

Both the pIL 11 and the pTA233 vectors were previously supplied on *E. coli* host strains on LB with ampicillin agar slants, where miniprep purification of each plasmid was performed as described in Section 3.3.2.7. pIL11 is a replication vector that multiplies in *E. coli* only. It is a 2845bp vector that possesses the *Hbt. salinarum* ferredoxin promoter (*fdx* promoter) for gene over-expression in *Hfx. volcanii* (Figure 3.2). As pIL11 lacks a *Lac Z* region, there is no possibility for blue/white screening. pTA233 is a shuttle vector that multiplies in both *Hfx. volcanii* and *E. coli* due to its dual origin of replications. It is a 7429bp vector and works well with clones up to 4kb (Figure 3.3). pTA233 vector derived from the pHV2 plasmid has a relative low copy number of 6 per cell (Charlebois et al. 1987), which is in contrast to the ColE1-based plasmids such as pUC19 with a copy number of over 100 per cell (Summers 1998).

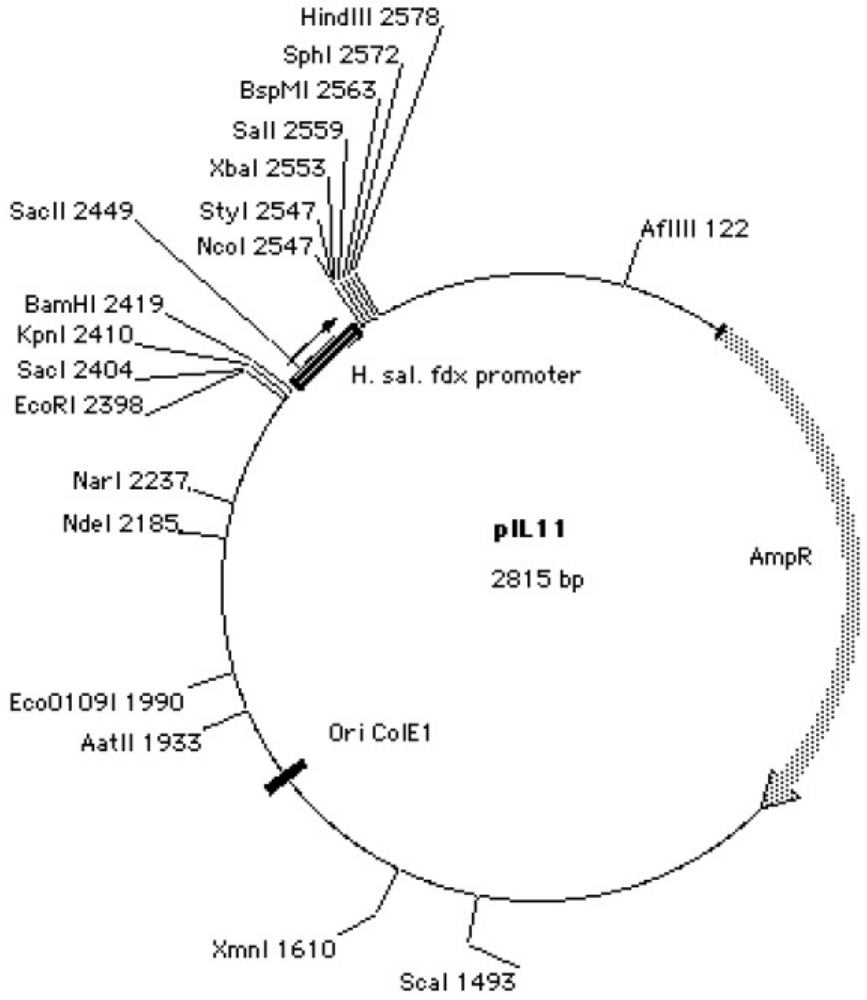


Figure 3.2: Circular map of pIL11 vector. The vector map of pIL11 includes the ampicillin/carbenicillin resistance gene (AmpR). This pUC19 based vector contains the Ori ColE1 replication origin from *E. coli*. *Fdx* promoter represents the ferredoxin promoter from *Hbt. Salinarum* for over expression in *Hfx. volcanii*. The *Nco*I site (C′CATGG) contains an ATG start codon for cloning of the target gene after the *fdx* promoter.

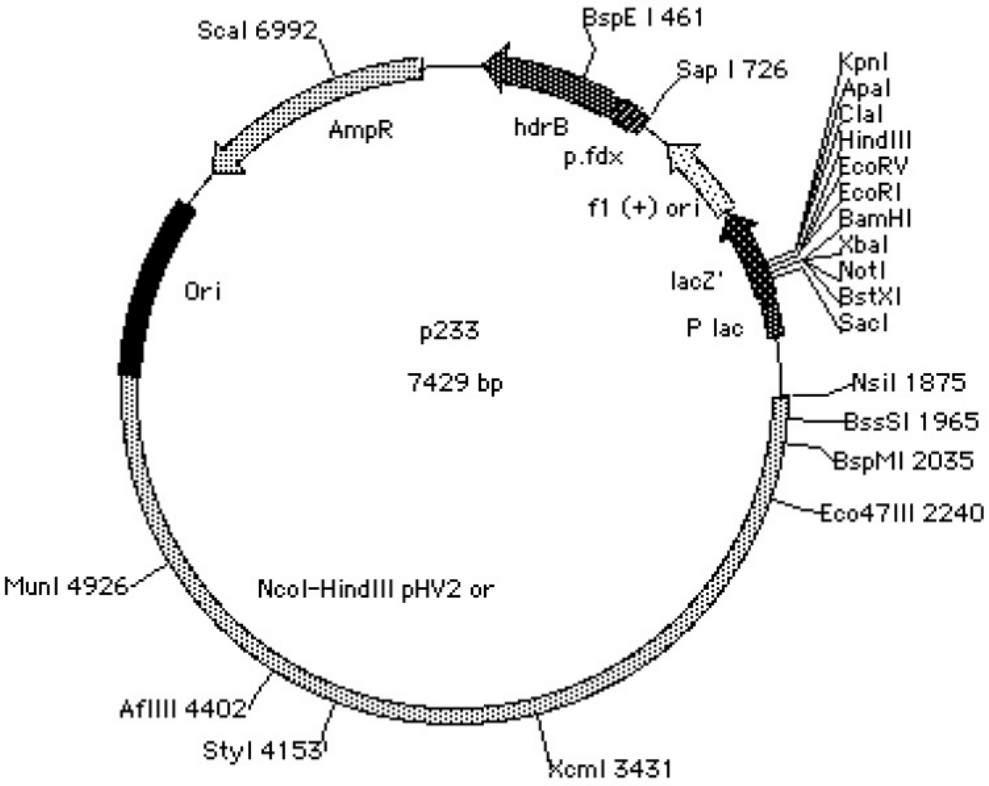


Figure 3.3: Circular map of pTA233 shuttle vector. pTA233 shuttle vector is derived from pTA192 vector. The map illustrates the LacZ region to allow for blue/white screening. The f1 (+) ori is the single-stranded DNA replication origin of the filamentous phage f1. The hdrB thymidine gene is a selectable marker that complements ΔhdrB *Hfx. volcanii* thymidine auxotrophy and is preceded by *fdx* promoter. This plasmid confers ampicillin/carbenicillin resistance gene (AmpR) selectable marker followed by the origin of replication (ori) of *E. coli*. The *Nco*I-*Hin*dIII pHV2 or contains the origin of replication for *Hfx. volcanii*.

3.3.3.2 Cloning procedure for *phr2*

The pIL 11 vector was used by previous lab members for the transformation of the 2-oxoacid dehydrogenase complex (OADHC) genes. One of the encoding genes, E3, was ligated into the pIL11 vector using *Xba*l and *Hin*dIII followed by transformation into JM109 competent cells. In this study, E3 was chosen as a negative control for the *phr* genes. After minipreps of the plasmids, the E3 insert along with the ferrodoxin promoter upstream of the gene were digested from the pIL11 vector using *Bam*HI and *Hin*dIII. This fragment was then ligated into respective site within the pTA233 shuttle vector, followed by transformation and minipreps. Due to the compatibility of the restriction enzymes used, the construct E3-pTA233 was digested using *Nco*I and *Hin*dIII to produce the linearized vector fragment without the insert; the *phr2* gene was then ligated into this linearized pTA233 vector, downstream of the already incorporated *fdx* promoter from pIL11 using the same restriction enzymes. This was followed by transformation into JM109 competent cells using the same method outlined in Section 3.3.2.6. The plasmids were purified following the procedure outlined in Section 3.3.2.7. The cloning procedure of E3 followed by *phr2* is outlined in Figure 3.4 below.

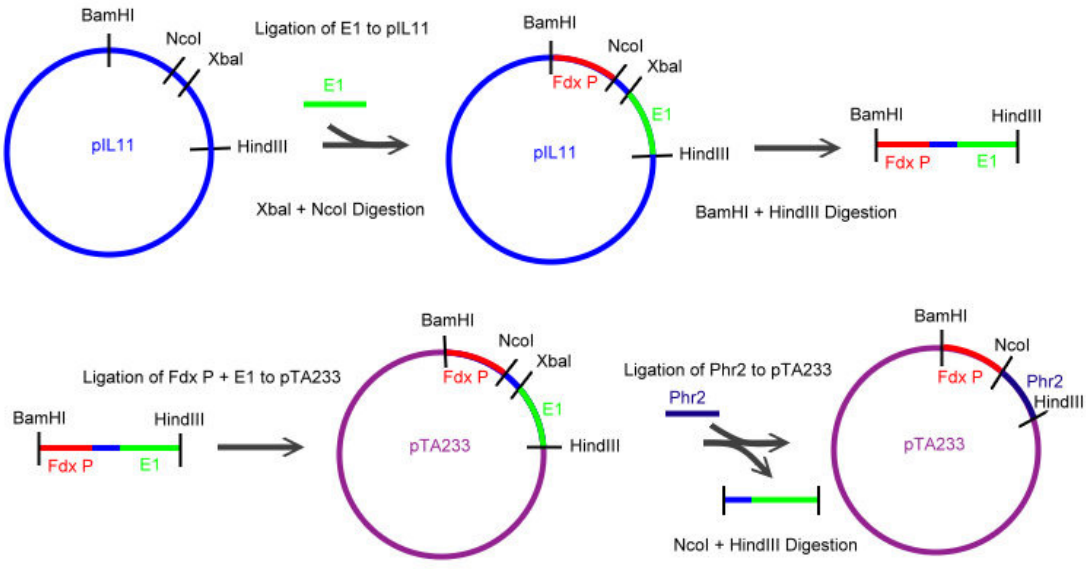


Figure 3.4: Schematic diagram of the *phr2* cloning procedure. The E3 gene, once cloned into the pIL11 vector, was digested using *Bam*HI and *Hin*dIII to incorporate the ferredoxin promoter (Fdx P), and the entire fragment was ligated into pTA233. As the gene of interest has an N-terminal *Nco*I site, a double digest was performed on pTA233 with *Nco*I and *Hin*dIII. The *phr2* gene with *Nco*I and *Hin*dIII overhangs, was then ligated into the linearized pTA233 vector.

3.3.3.3 Cloning procedure for *phr1*

The cloning of the *phr1* differed from that of *phr2* due to the difference in restriction enzyme used. The excised *phr1* fragment was cloned directly into pIL11, which was also digested with the same endonucleases (*Xba*I and *Hin*dIII). The *fdx* promoter, along with *phr1* were excised and ligated into the pTA233 vector, followed by transformation into JM109 competent cells as outlined in Section 2.3.2.6. The route for the *phr1* cloning procedure is shown in Figure 3.5 below.

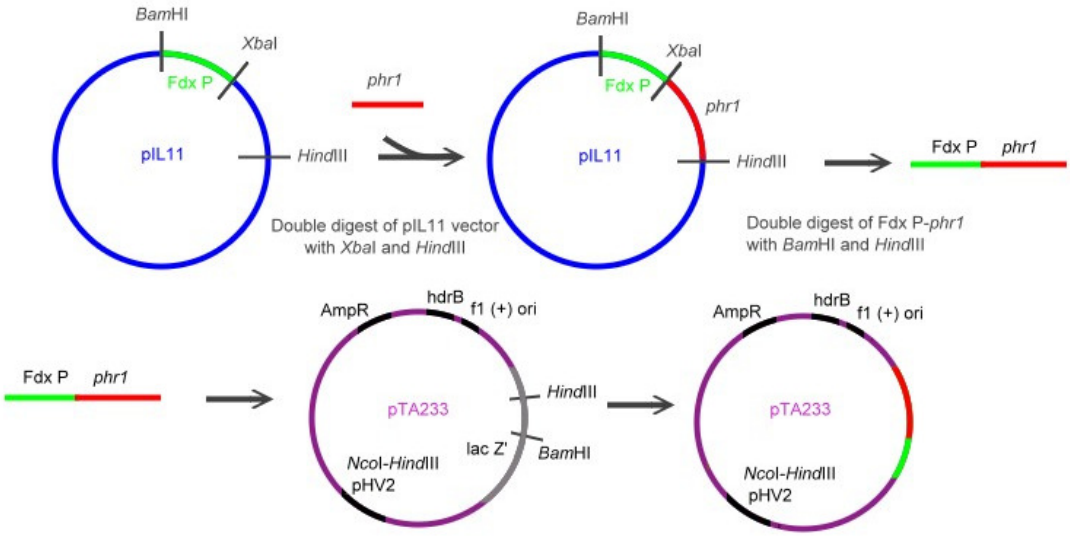


Figure 3.5: Schematic diagram of the *phr1* cloning procedure. The *phr1* gene was ligated into the digested pIL11 vector with *Xba*I and *Hin*dIII. Along with the fdx promoter (Fdx P), *phr1* was digested out of pIL11 with *Bam*HI and *Hin*dIII, which was subsequently ligated into pTA233 using the same restriction enzymes.

3.3.4 Transformation of competent *dam-/dcm- E. coli* cells

Following the High Efficiency Transformation Protocol as outlined in the New England Biolabs leaflet, 3µl of the purified plasmid containing either *phr2* or *phr1* was added to the *dam-/dcm- E. coli* cell mixture and placed on ice for 30 min. The cells were heat shocked at exactly 42°C for 30 sec and placed on ice for 5 min. 250µl of SOC medium was added to the mixture, placed in a shaking incubator (180 rpm) at 37°C for 60 min. 100µl of diluted cell mixture was plated onto ampicillin LB plates and incubated at 37°C for one day. Single colonies were inoculated into 10ml LB containing 10µl carbenicillin, followed by incubation in a 37°C shaking incubator (180 rpm). The plasmids were purified following the procedure as outlined in Section 2.2.3.7.

3.3.5 Growth of *Hfx. volcanii* H98 expression strain

The *Hfx. volcanii* host strain H98 is of (Δ*pyrE2* Δ*hdrB*) genotype and/or phenotype. A deletion in both of these genes results in uracil and thymidine auxotrophy in rich medium (Hv-YPC). The H98 expression strain was grown in a rich medium (Hv-YPC) composed of two separate solutions. The first is a 30% salt water (SW) solution consisting of 24% (w/v) NaCl, 3% (w/v) $MgCl_2 \cdot 6H_2O$, 3.5% (w/v) $MgSO_4 \cdot 7H_2O$, 0.7% (w/v) KCl, and 20 ml of 1M Tris-HCl, pH 7.5, all dissolved in warm water to a final volume of one litre which was filter sterilized. The second part of the medium is 10 x YPC consisting of 8.5g yeast extract (Difco), 1.7g peptone (Oxoid) and 1.7g Casamino Acids (Difco), all dissolved in distilled water to a final volume of 170ml and adjusted to pH 7.5 with 1M KOH. To make up liquid Hv-YPC medium, 200ml of 30% salt water was mixed with 33ml of 10X YPC and added into a clean 500ml Duran bottle. A final volume of 333ml was made up with distilled water. The medium was autoclaved, and both 0.5M $CaCl_2$ and filter sterilized thymidine (40µg/ml) at final concentrations were added upon cooling. To make solid Hv-YPC medium, 100ml of distilled water and 200ml of 30% salt water was mixed with 5.0g of agar. The contents were dissolved in the microwave with continuous mixing in between. 30ml of 10X YPC was added to the contents above and the solution was autoclaved. After cooling to about 57 °C, 2ml of 0.5M $CaCl_2$ and filter sterilized thymidine (40µg/ml) were added. The plates were poured thick, sealed in plastic and stored in a dark cupboard.

3.3.6 Transformation of *Hfx. volcanii* H98 expression strain

Unmethylated pTA233 with the cloned inserts were transformed into *Hfx. volcanii* H98 expression strain. This transformation protocol was carried out using polyethylene glycol (PEG). A freshly inoculated culture (10ml Hv-YPC with thymidine) was grown overnight to late log phase at OD_{650} of approximately 0.8. This was harvested at 4,500 x g for 8 min at room temperature. The solutions shown in Table 3.5 were prepared in advance for the transformation procedure.

Table 3.5: Buffers for transforming H98 strain

Buffered sterile spheroplasting solution	
NaCl	14.61g
KCl	0.5g
Tris-HCl (pH8.5)	12.5ml
Sucrose	37.5g
Add distilled water to 250ml	

Unbuffered sterile spheroplasting solution	
NaCl	5.84g
KCl	0.2g
Sucrose	15g
Add distilled water to 100ml	
Adjust pH to 7.5 with 1M NaOH	

Polyethylene glycol 600 at 60%	
PEG 600	480µl
Unbuffered spheroplasting solution	320µl

Sterile spheroplasting dilution solution	
30% salt water	76ml
Sucrose	15g
0.5M $CaCl_2$	0.75ml
Add distilled water to 100ml	

Sterile regeneration solution	
30% salt water	150ml
10x YPC	25ml
Sucrose	37.5g
0.5M CaCl2	1.5ml
Add distilled water to 250ml	

Sterile transformation dilution solution	
30% salt water	150ml
Sucrose	37.5g
CaCl2	1.5ml
Add distilled water to 250ml	

The pellet was washed in 2ml sterile buffered spheroplasting solution, transferred to a 2ml round bottomed tube and pelleted as above. The pellet was then gently re-suspended in 600µl of buffered spheroplasting solution. For each transformation, 200µl of cells were transferred to a clean 2ml round bottom tube. 20µl of 0.5M EDTA (pH 8.0) was added to the side of the tube, and the contents were mixed by inverting, followed by incubation at room temperature for 10 min to let the

spheroplasts form. Meanwhile, DNA samples were set up as follows: 10µl dam-DNA (or 10ul of sterile water as control), 15µl unbuffered spheroplasting solution, and 5µl 0.5M EDTA (pH 8.0). After 10 min, the DNA was added to the spheroplasts and incubated for 5 min. After 5 min, 250µl of 60% PEG was added to each transformation. The tube was shaken horizontally for about 10 times, and left at room temperature for 30 min. 1.5ml spheroplast dilution solution was added to the tube, inverted to mix, and left at room temperature for 2 min. The cells were harvested at 4,500 x g for 8 min at room temperature, and the supernatants removed. 1ml regeneration solution and 60µg/ml thymidine were added to the pellet, and the whole pellet was transferred to a 4ml Bijou tube. The tube was left undisturbed at 45°C for 1.5-2 hr for regeneration to occur. At the end of the incubation period, the pellet was re-suspended by tapping the side of the Bijou, and return to 45°C on shaking at 150rpm for 3-4 hr. After incubation, the cells were transferred to a clean 2ml round bottom tube, and pelleted as before. The supernatant was removed, and the pellet was resuspended gently in 1ml transformant dilution solution. 100µl of cells was plated onto Hv-YPC plates with and without thymidine using the dilutions 10^0, 10^1, and 10^2 in transformant dilution solution, and incubated at 45°C and sealed in plastic.

After 5 days when the colonies had grown, the colonies were picked from each transformation and placed into 50ml flasks containing 20ml Hv-YPC medium. The cells were incubated at 45°C on shaking at 180 rpm. When the OD_{600} reached higher than 1.2, 5ml of the culture was removed for later use, and the rest of the cells were harvested and stored in either the -20°C or -80°C freezers as glycerol stocks.

3.3.7 Analysis of transformants

5ml of the cell culture was spun down and used to prepare minipreps of plasmid DNA as mentioned in Section 3.3.2.7. *Nco*I, *Xba*I, and *Hin*dIII were used for the double digestion of the plasmids to determine if the transformation was successful. Cell extracts and cell debris were examined by SDS-PAGE on a 10% (w/v) polyacrylamide gel (Laemmli 1970) with standard buffers (Sambrook and Russell 2001).

3.3.7.1 Preparation of cell extracts for SDS-PAGE

10ml cell culture was inoculated, and 1ml was used to measure the absorbance at OD_{600}. The weights of the tubes were also measured. The 1ml was transferred to 2ml tubes and harvested at 8,000 rpm for 3 min. The supernatants were removed, and the cell paste was re-suspended in extracting buffer (0.2g cell paste per 1ml buffer). For example, 0.4g of cell paste was re-suspended in 2ml buffer. Using the 150-W Ultrasonic Disintegrator (MSE Scientific Instruments), the extracts were then subjected to three 30s bursts of sonication (amplitude = 14 micron) in an ice bath with 60s of cooling between each burst. The cell debris was removed by centrifugation at 16,000 x g for 10 min, and the supernatants were poured into clean 1.5ml tubes. Equal amounts of 2X SDS loading buffer was added to the supernatants according to the OD readings taken earlier. The cell debris was re-suspended in 50-100µl of sterile water, and 10µl of debris was mixed with 10µl of 2X SDS loading buffer. The samples were placed in the -20°C for further use.

3.3.7.2 SDS-PAGE

10µl of the sample was run on a SDS-PAGE gel alongside the BioRad protein marker which contained the following proteins: myosin (M_r 210Kda), β-galactosidase (116.3), phosphorylase b (97.4), bovine serum albumin (66.2), ovalbumin (45.0), carbonic anhydrase (31.0) and trypsin inhibitor (21.5). The protein marker was prepared by diluting 1 in 10 with sterile water, followed by the addition of 2X SDS loading buffer in a 1:1 ratio, and 2µl of 5% β mercaptoethanol. The mixture was boiled for 3 min at 95°C. The gel was run at 10mA for 2 hr, followed by staining in Coomassie Brilliant Blue R250 and de-staining in (700ml H20, 200ml Methanol, and 100ml acetic acid) overnight.

3.4 Results

3.4.1Cloning of the *phr2* and *phr1* genes into pGEM®-T Easy Vector

3.4.1.1 PCR amplification of the *phr* genes

Due to the high GC content of the *Hfx. volcanii* genome and the repetitive nature of the nucleotide sequences, it was difficult to design a set of primers that only annealed to the desired sequence. The first set of *phr2* primers failed to produce any

products and there appeared to be smears in the control. After running a second set of gels where no smears were found in the control, it appeared that the problem may have been attributed to primer specificity. The next checkpoint was to adjust the PCR conditions by increasing the annealing temperature for greater specificity. A touchdown PCR was used in which a range of temperatures from 56°C to 64°C were tested. This is a method by which non-specific binding is decreased by first setting a very high annealing temperature, which at every subsequent cycle, is decreased by 1°C. As primers bind specifically at a temperature just below their melting temperatures, the first sequences amplified would be of greatest primer specificity, meaning that it is most likely the sequence of interest. These sequences will still bind at lower temperatures, and the non-specific sequences would be bound comparatively less by the primers (Don et al. 1991). Although touchdown PCR gave a higher yield in the wanted product, these primers were still not specific enough as there was still unwanted background. It was decided to re-design a new set of primers that included a higher Tm and specificity (Table 3.6). The highlighted sequence indicated the *Nco*I restriction site that incorporated the two base pair changes to the *phr2* gene sequence.

Table 3.6: Primer sequences for amplifying *phr2*

Primer	Sequence 5′-3′	Tm °C
Hfx F	GCGCCCATGGAACTGCACTGGCACCG	74.9
Hfx R	CCAAGCTTGACCGAGGGCAAAAAACAGCAG	75.3

The new primers were blasted against the whole genome to see whether non-specific binding occurred. The forward primer did not contain any non-specific binding while the reverse primer annealed to two other sequences within the genome, although the E values were too large to have significant effects. The annealing temperature was increased to 70°C to increase specificity. In this PCR, Vent$_R$®, *Taq*, and Phusion™ polymerases were all tested to determine the one that gave the best yields. 5µl of the PCR products were visualized on a 1.0% (w/v) agarose gel for analysis. Vent$_R$® polymerase gave the highest yield in the product of interest, with some non specific bands at comparatively lower concentrations. *Taq* polymerase yielded less PCR product in general when compared with Vent$_R$® polymerase (Fig 3.6a). Phusion™ polymerase failed to amplify any DNA (Fig 3.6b). The size of the

band of interest was calculated to be 1554bp, 99bp larger than the expected *phr2* size of 1455bp. The remainder of the $Vent_R^®$ amplified PCR products were run on a 1.0% gel (w/v), and the band of the correct size was carefully excised from the gel and purified using the QIAquick Gel Extraction kit and purification method in Section 3.3.2.3. As the PCR conditions were optimized for *phr2* already, the amplification of *phr1* was performed with ease. The only change to the thermocycling condition is a decrease in annealing temperature to 56°C to accommodate the lower Tm of the primers. The PCR fragment produced was clean with the expected size; the band was extracted and purified as above (Fig 3.7). The primer sequences are shown in Table 3.7.

Table 3.7: Primer sequences for amplifying *phr1*

Primer	Sequence 5′-3′	Tm °C
phr1forward	TCTAGAACGATGTCCGAGCCAACTTC	62.5
phr1reverse	AAGCTTGCATCTAATCGGAGAAATCG	61.7

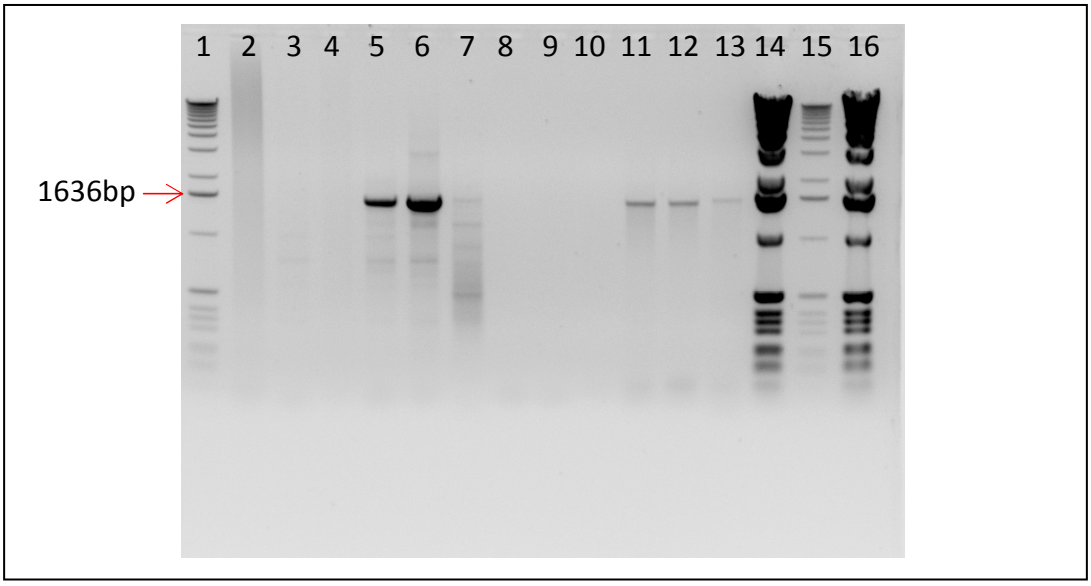


Figure 3.6a: Gel electrophoresis analysis of PCR amplified *phr2* gene products using $Vent_R^®$ and *Taq* polymerases. Lanes **2** to **7** were amplified using $Vent_R^®$, while lanes **8** to **13** were amplified using *Taq*. From left to right: lane **1**, **14**-**15** = 1 kb DNA ladder; **2** = negative control; **3** = HfxR primer only; **4** = HfxF primer only; **5**&**6** = HfxF & HfxR primers; **7** = old primers (forward and reverse); **8** = negative control; **9** = HfxR prime only; **10** = HfxF primer only; **11**&**12**= HfxF & HfxR primers; **13** = old primers (forward and reverse)

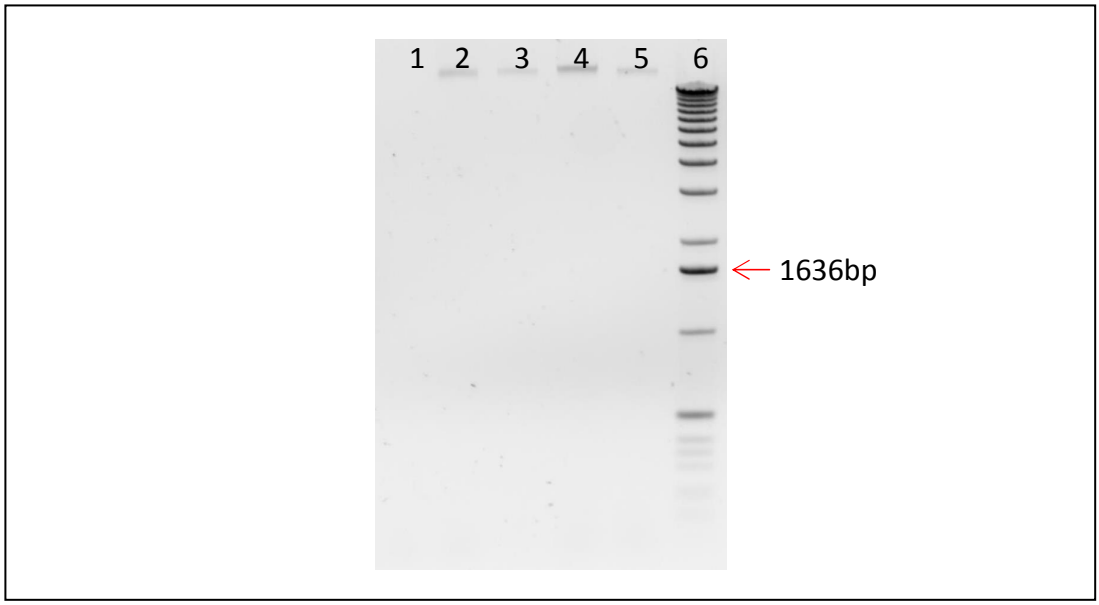


Figure 3.6b: Gel electrophoresis analysis of PCR amplified *phr2* gene products using Phusion™ polymerase. From left to right: lane **1** = negative control; **2** = HfxR primer only; **3** = HfxF primer only; **4** = HfxF & HfxR with GC buffer; **5** = HfxF & Hfx R with HF buffer; **6** = 1 kb DNA ladder

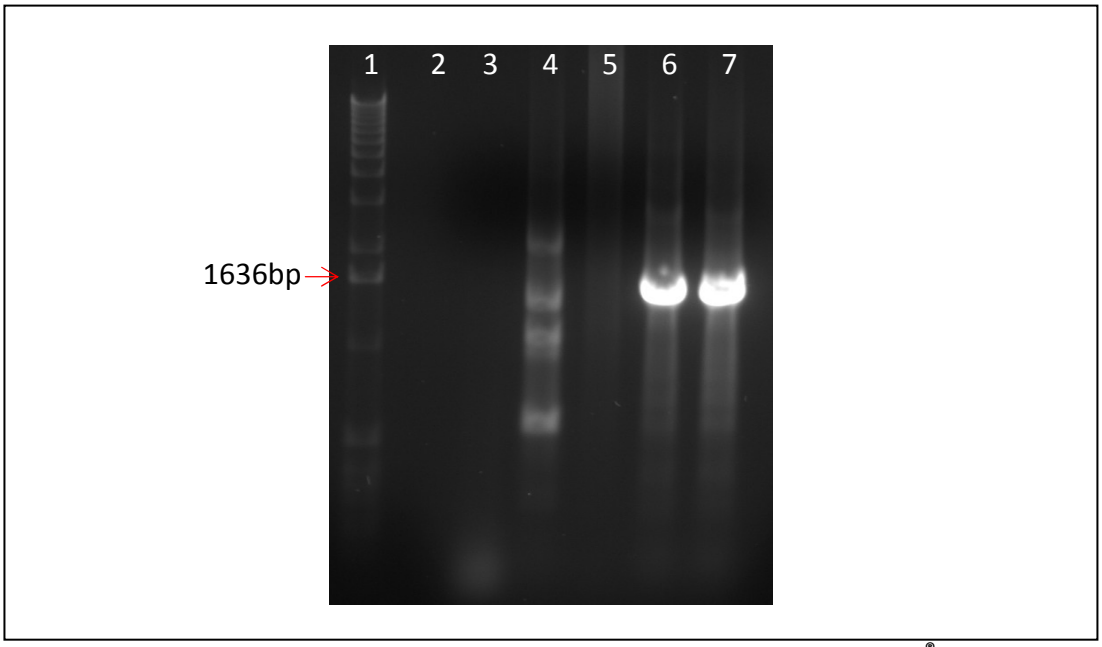
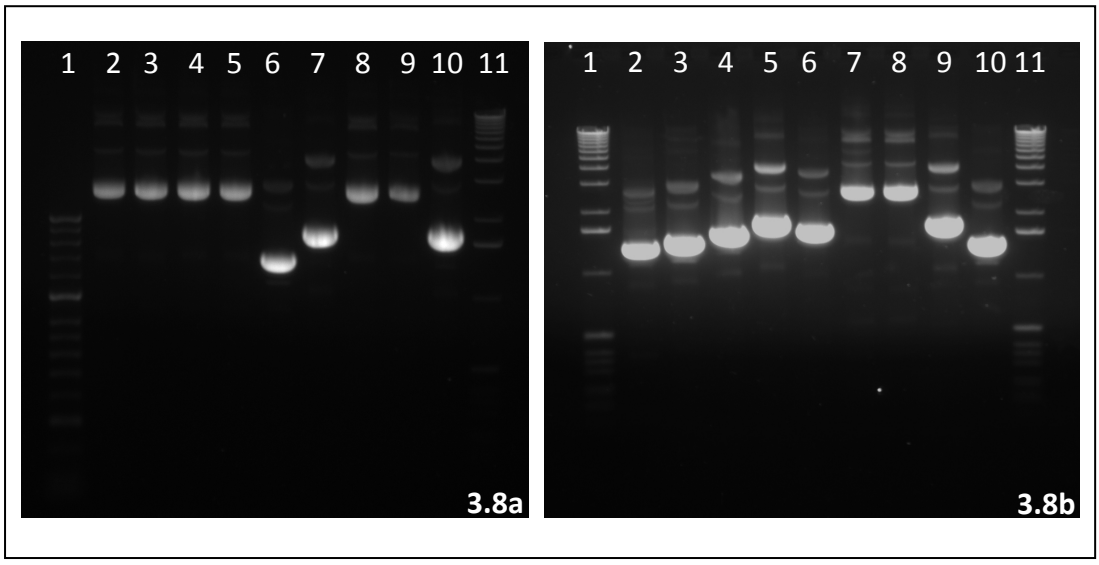


Fig 3.7 Gel electrophoresis analysis of PCR amplified *phr1* using Vent$_R$® Polymerase. From left to right: Lane **1** = 1 kb DNA ladder; **2** = blank**; 3** = negative control; **4** = phr1forward primer only; **5** = phr1reverse primer only; **6**&**7** = phr1forward and phr1reverse primers

3.4.1.2 A-tailing, ligation and cloning

A-tailing was performed on the isolated fragment to ensure that the DNA could be ligated into the pGEM®-T Easy Vector. The *phr2*-containing plasmids were then transformed into highly competent *E. coli* cells for blue/white screening. The cells were grown on LB ampicillin plates, and 18 white and blue colonies were selected and grown in liquid culture overnight. The plasmids from the cells were purified and visualized on an agarose gel to determine which plasmids contained the correct insert. Plasmids in lanes 2, 3, 4, 5, 8, 9 from Figure 3.8a and lanes 7 and 8 from Figure 3.8b appeared to be bigger in size compared to the rest of the plasmids at 3000bp; therefore, these plasmids may contain the insert. When checked with the master plate, these plasmids correspond to the white colonies on the plate. Plasmids in all other lanes were from blue colonies. Before the plasmids were sent for sequencing, the results were confirmed again with single and double digests to check for the success of transformation (Figure 3.9). Purification of plasmid preps for *phr1*-pGEM®-T (figure for plasmid preps alone not shown) and restriction digest was performed as shown in Figure 3.10.



Figures 3.8a and 3.8b: Gel electrophoresis analysis of plasmids of *phr2*-pGEM®-T construct. **3.8a -** Lanes from left to right: 1 = 100bp DNA ladder; 2-10 = purified plasmids; 11 = 1 kb DNA ladder. **3.8b -** Lanes from left to right: 1 = 1 kb DNA ladder; 2-10 = purified plasmids; 11 = 1 kb DNA ladder

3.4.1.3 Restriction digest

A single digest was performed with *Hin*dIII on all potential *phr2*-containing plasmids, and the samples were run on an agarose gel for analysis. The plasmids with the correct size were subsequently digested with *Nco*I to confirm the presence of the product at ~1500bp (Figure 3.9). In Figure 3.9, plasmids corresponding to lanes 2 to 9 were digested to show a top band at ~3000bp corresponding to the size of the vector used, and a lower band at approximately 1500bp, which is most likely the *phr2* gene. Plasmids in lanes 10 and 11 did not contain an insert; therefore, the digest only resulted in pGEM®-T vector alone cut with *Nco*I only, as the *Hin*dIII restriction site does not exist within the vector sequence. The lower bands at ~1500bp were excised from the agarose gel and purified. To check for potential *phr1*-containing plasmids, both single and double digests were performed on all 10 purified plasmids with *Xba*I and *Hin*dIII (Figure 3.10). Plasmids corresponding to lanes 10 and 12 (white colonies on plate) contained the vector at ~3000bp and insert of the right size at ~1500bp. The 1500bp fragments were excised from the agarose gel and purified.

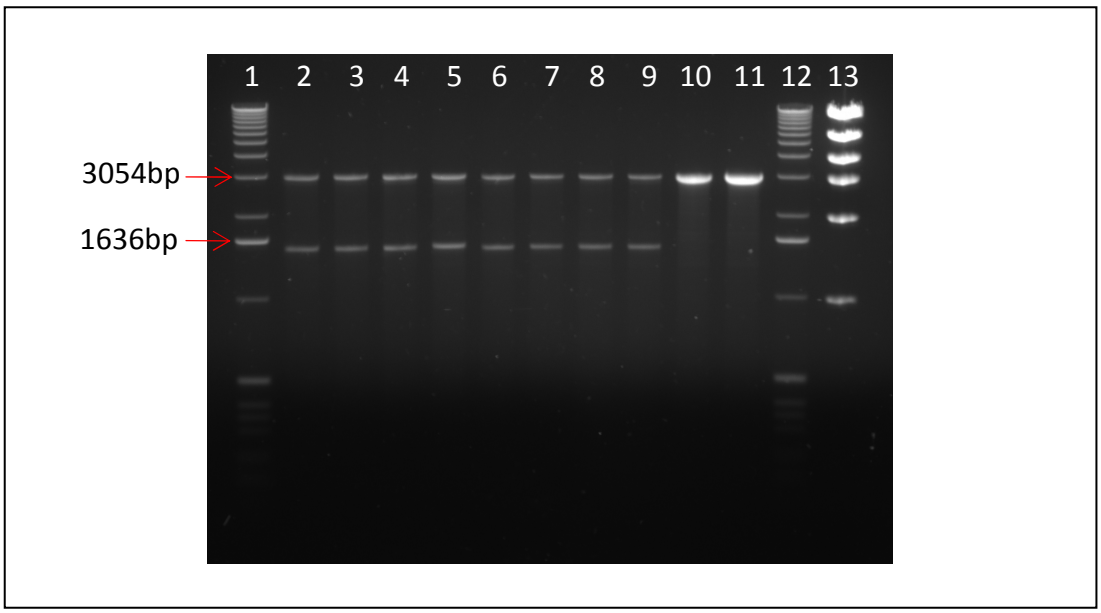


Figure 3.9: **Gel electrophoresis analysis of double digests of all plasmids containing the *phr2* insert.** Lanes from left to right: **1**&**12**= 1kb DNA ladder; **2**-**9** = plasmids digested with *Hin*dIII and *Nco*I; **10**&**11** = digested plasmids without insert; **13** = DNA mass ladder

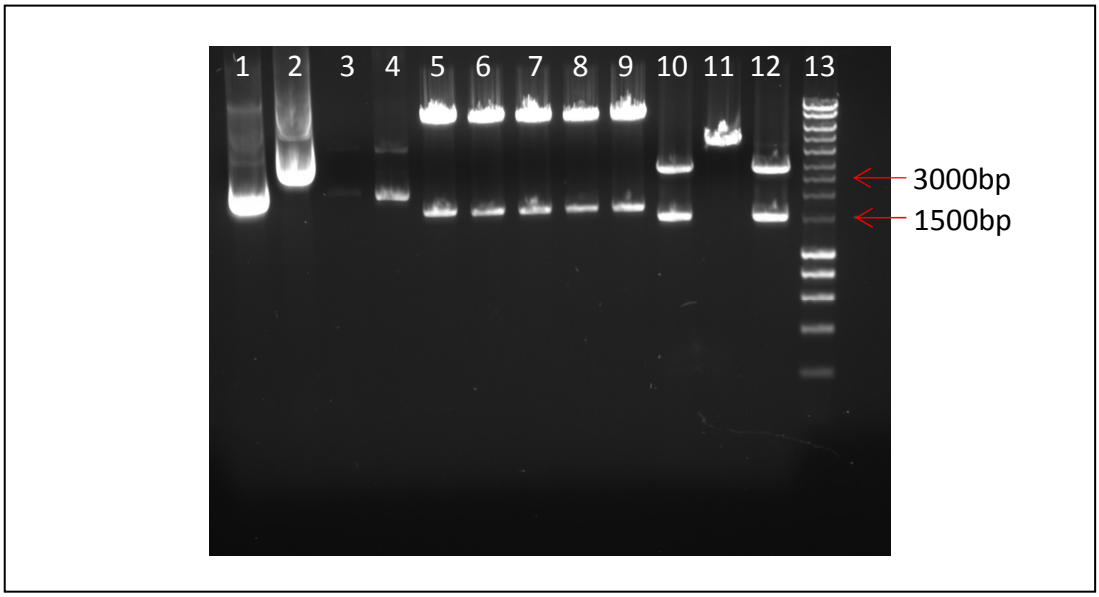


Figure 3.10: Gel electrophoresis analysis of double digest of all plasmids containing the *phr1* insert. Lanes from left to right: **1** = uncut blue colony (background control); **2** = positive control (uncut white colony); **3**&**4** = blue colony cut with *Xba*I and *Hin*dIII; **5**-**12** = white colony cut with *Xba*I and *Hin*dIII; **13** = 1 kb DNA ladder

3.4.1.4 Sequencing

All 8 clones corresponding to lanes 2 to 9 from Figure 3.9 were shown to contain the *phr2* gene and these plasmids were sequenced and aligned to the putative *phr2* sequence to check for sequence identity and to determine the best clone for subsequent cloning experiments. *Nco*I was chosen as the restriction site for the forward primer because it contains the ATG start codon sequence which ensured that the *phr2* gene would be present after insertion into the pGEM®-T vector. However, including this site in the forward primer involved two base pair mismatches on both sides of the start codon. This resulted in a second residue change from a serine to an alanine. As both residues are uncharged and are of similar sizes, this should not affect the subsequent expression of the *phr2*-encoded protein. All 8 clones contained the base pair changes. Sequencing results also revealed that the sequence of clone 1 (lane 2) was identical to the putative *phr2* sequence, with the exception for one base pair change resulting in an amino acid change from alanine to a valine. Whilst valine has a bigger side chain compared to alanine, both amino acids are hydrophobic, non-polar and neutral. This clone was chosen because its sequence was very clean compared to the sequences of the other 7 clones which did not have

this base pair change. For *phr1*, sequencing was performed for clones corresponding to lanes 10 and 12 (colonies 8 and 10 on plate) from Figure 3.10, both of which revealed 100% sequence identity to the putative *Hfx. volcanii phr1* sequence. Clone 8 was chosen for subsequent cloning work.

3.4.2 Cloning of *phr* genes into pIL11 and pTA233 vectors

Both the pIL11 and pTA233 vectors were initially supplied in *E. coli* host strains on LB ampicillin agar slants. Both pIL11 and pTA233 vectors had previously been used for the cloning of the dihydrolipoamide dehydrogenase (E3) gene from *Hfx. volcanii*. E3 was digested from pGEM®-T using *Xba*I and *Hin*dIII and ligated into the pIL11 vector cut with the same enzymes, followed by transformation into JM109 competent cells where the resulting plasmids were purified. The *fdx* promoter, along with E3, were digested from pIL11 using *Hin*dIII and *Bam*HI, where the whole fragment was then ligated into the pTA333 vector and transformed into competent cells (AL-Mailem 2006). For convenience, the purified plasmid preps containing E3 were used for the cloning of the *phr2* fragment. This was done by a double digest of the *fdx*-E3-pTA233 construct with *Nco*I and *Hin*dIII to remove the E3 fragment, leaving the intact *fdx* promoter in the linearized vector. The *phr2* fragment digested with the same enzymes was ligated into the linearized pTA233 vector, followed by transformation in competent cells and purification of the plasmids. The *phr2* gene was now preceded by the *fdx* promoter in pTA233.

The plasmids were subjected to double digests with *Nco*I and *Hin*dIII to ascertain the presence of the *phr2* fragment. The digested plasmids were run on a 1% agarose gel as shown in Figure 3.11a and 3.11b. Digested plasmids corresponding to lanes 2 to 7 (clones 1-6) and lanes 10 to 12 (clones 9-11) in Figure 3.11a contained the insert of the right size at ~1500bp. The top band (~7000bp) corresponded to the size of the pTA233 vector with a size of 7.4kb. In Figure 3.11b, digested plasmids corresponding to lanes 2, 3, 6, 8 and 12 (clones 1, 2, 5, 7 and 10) also appeared to contain two fragments of the right sizes as well. The undigested plasmids, designated the *fdx-phr2*-pTA233 construct, were stored in the -20°C and as glycerol stocks in -80°C freezers for future use. Plasmids corresponding to lanes 2 and 3 from Figure 3.11a were used for subsequent experiments.

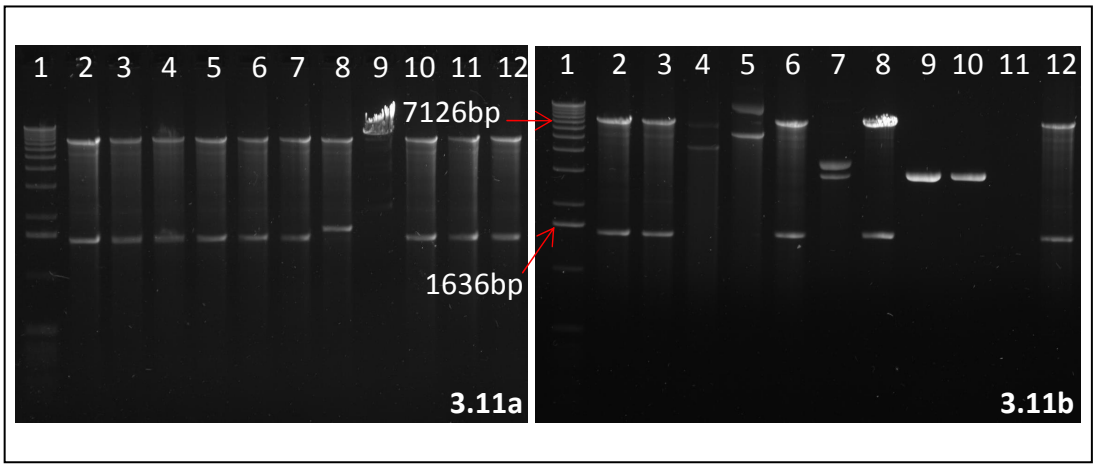


Figure 3.11a and Figure 3.11b: Gel electrophoresis analysis of double digest of *fdx-phr2*-pTA233 construct. **3.11a -** lanes from left to right: **1** = 1 kb DNA ladder; **2**-**12** = double digested plasmids with *Hin*dIII and *Nco*I. **3.11b -** lanes from left to right: **1** = 1 kb DNA ladder; **2**-**10**, **12** = double digested plasmids with *Hin*dIII and *Nco*I; **11** = blank

The cloning of *phr1* followed a more conventional route where it was first digested from pGEM®-T using *Xba*I and *Hin*dIII and the fragment gel purified. The *phr1* fragment was ligated into the pIL11 vector previously digested with the same enzymes. Transformation of the construct into competent cells was performed and the plasmids were purified. A double digest using *Xba*I and *Hin*dIII was performed to check for successful ligations (Figure 3.12a). All plasmids amplified fragments at ~1500bp and 2800bp, corresponding to the sizes of *phr*1 and pIL11, respectively. The purified plasmids were also double digested with *Bam*HI (upstream of the *fdx* promoter) and *Hin*dIII (to include the *fdx* promoter with the *phr1* gene), and all four plasmids contained the insert (Figure 3.12b). This *fdx-phr1* fragment was gel extracted and purified, and ligated into pTA233 shuttle vector previously digested with *Bam*HI and *Hin*dIII. Transformation in competent cells was performed, followed by purification of the plasmids and a double digest with *Bam*HI and *Hin*dIII to ascertain the presence of the *fdx-phr1* fragment in pTA233. Figure 3.12c shows that only 1 plasmid corresponding to lane 4 was successfully digested to contain two fragments of the right sizes for *phr1* (~1500bp) and pTA233 (~7000bp). This plasmid was stored in -20°C and as glycerol stocks in -80°C freezers for subsequent transformations.

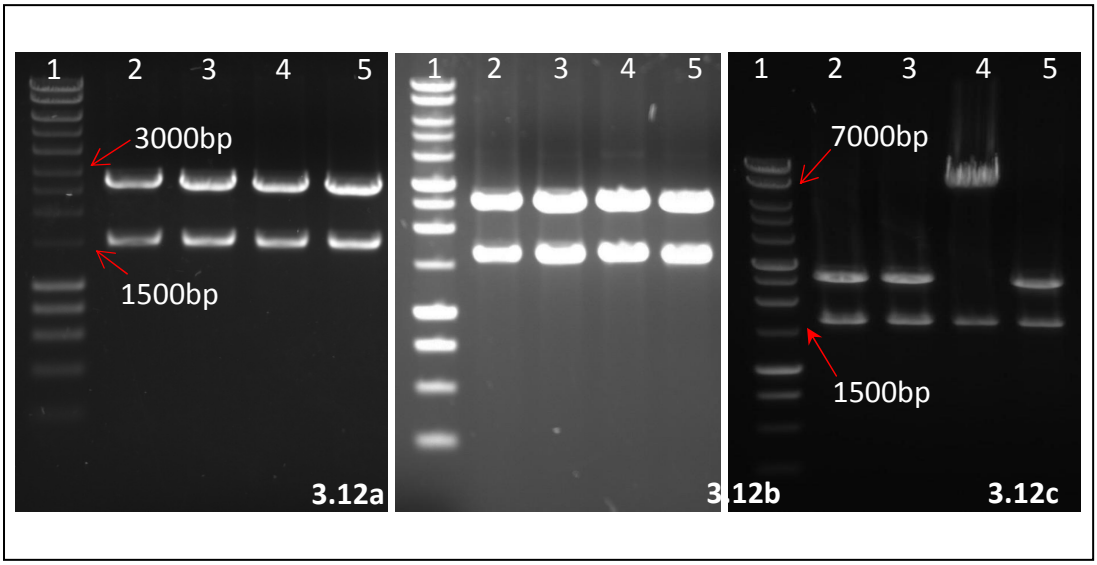


Figure 3.12a, 3.12b and 3.12c: Gel electrophoresis analysis of restriction digests
Figures 3.12a: Double digest of *phr1*-pIL11 construct with *Xba*I and *Hin*dIII. Lanes from left to right: **1** = 1kb DNA ladder; **2**&**4** = Colony 8, **3**&**5** = Colony 10
Figure 3.12b: Double digest of *phr1*-pIL11 construct with *Bam*HI and *Hin*dIII. Lanes from left to right: **1** = 1kb DNA ladder; **2**&**4** = Colony 8; **3**&**5** = Colony 10
Figure 3.12c: Double digest of *phr1*-pTA233 construct with *Bam*HI and *Hin*dIII. Lanes from left to right: **1** = 1kb DNA ladder; **2**&**4** = Colony 8; **3**&**5** = Colony 10

3.4.2.1 Sequencing

To ascertain that the *fdx* promoter was incorporated along with the *phr2* gene, sequencing was performed on the transformed *fdx-phr2*-pTA233 construct. Both the sequences of the promoter and the *phr2* gene were identified from the sequencing results obtained, indicating that the cloning was successful. This was not done for *phr1* as double digests with *Bam*HI and *Hin*dIII revealed two bands of the right sizes, indicating that the *fdx* promoter (with an upstream *Bam*HI site) was indeed incorporated within the construct. The *Bam*HI site needed to have been incorporated into the construct in order for the digest to work.

3.4.3 Transformation of *dam-/dcm- E. coli* cells

The DNA of the clones that were transformed using JM109 *E. coli* competent cells will be methylated, targeting it for degradation upon transformation into *Hfx. volcanii* H98 strain (Holmes et al. 1991). To resolve this problem, the plasmids were transformed into a non-methylating (*dam-/dcm-*) strain for the efficient transformation of *Hfx. volcanii*. The newly unmethylated *fdx-phr2*-pTA233 plasmids were subjected to double digest with *Nco*I and *Hin*dIII to determine

whether the transformation was successful. Figure 3.13 showed that 6 out of 7 clones contained the *fdx-phr2*-pTA233 construct; therefore, clone 1 (lane 2) and clone 2 (lane 3) were chosen for the transformation of the H98 native strain. 4 unmethylated *fdx-phr1*-pTA233 plasmids were subjected to only a single digest with *Bam*HI (Figure 3.13b), which showed that the construct was approximately 9000bp in size, corresponding to the size of the whole construct including *phr1* (~1500bp) and pTA233 (~7500bp). Plasmid 1 in lane 2 was chosen for the transformation of the H98 native strain.

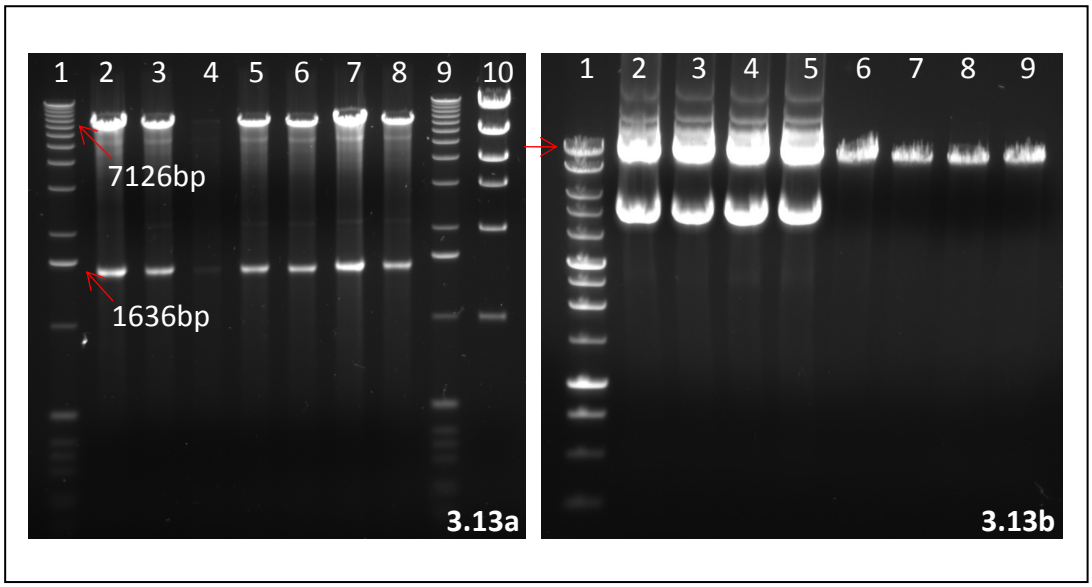
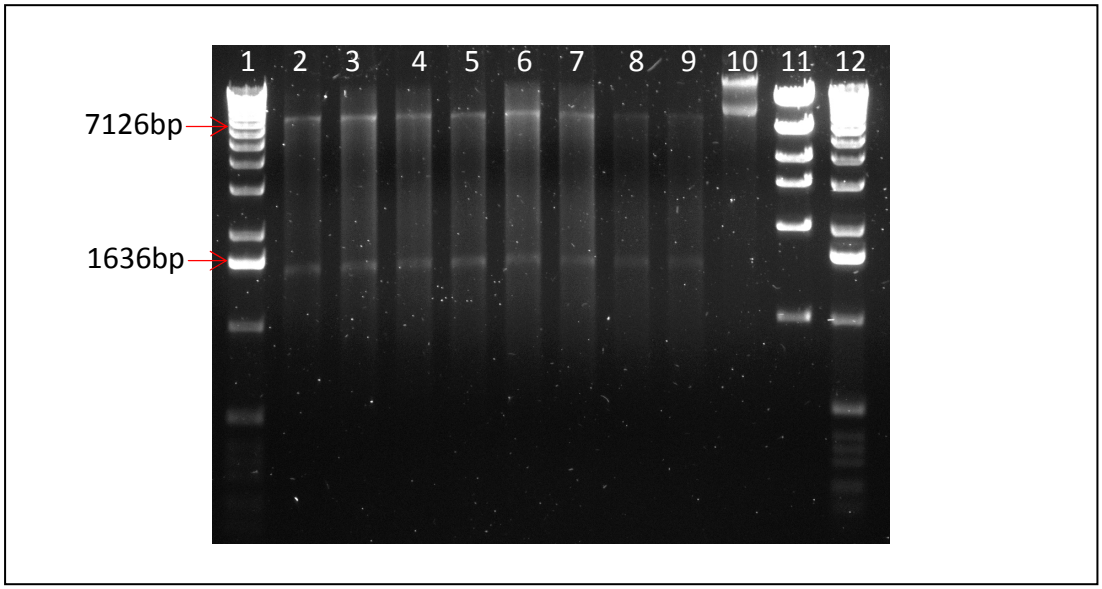


Figure 3.13a and 3.13b: Gel electrophoresis analysis of restriction digests and plasmid preps
Figure 3.13a: Double digest of de-methylated *phr2*-pTA233 construct (*Nco*I and *Hin*dIII) From left to right: Lane **1**&**9** = 1kb DNA ladder; **2**-**8** = digested plasmids; **10** = DNA mass ladder
Figure 3.13b: Plasmid preps and single digest (*Bam*HI) of de-methylated *phr1*-pTA233 construct (arrow indicated 9000bp band of the kb ladder)**.** From left to right: Lane 1 = 1kb DNA ladder; 2-5 = plasmid preps; 6-9 = digested plasmids

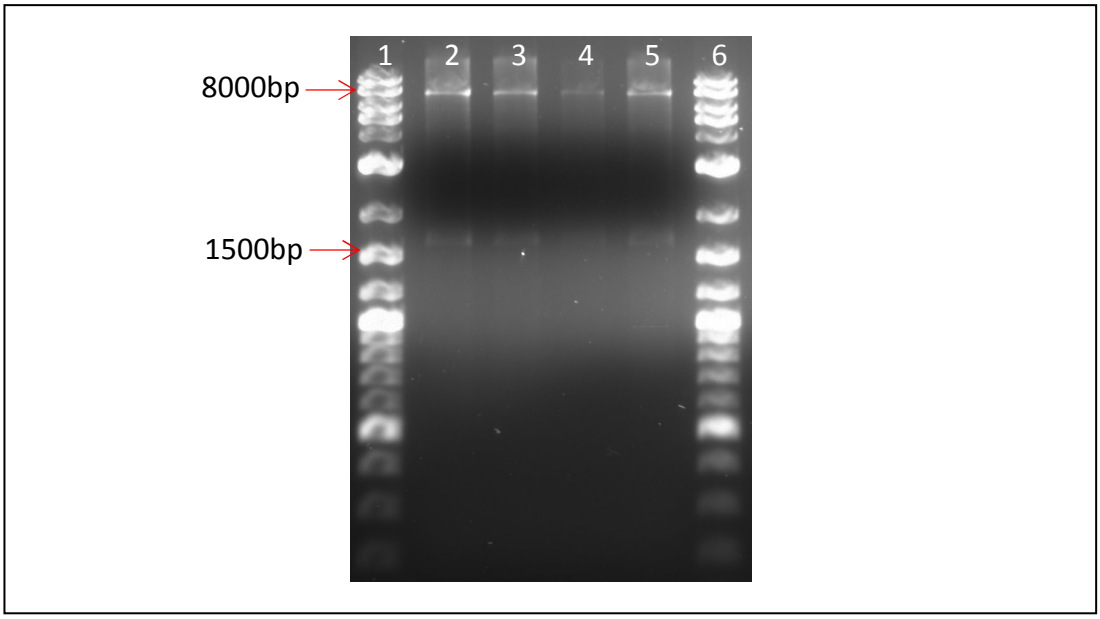
3.4.4 Transformation of *Hfx. volcanii* H98 expression strain

When the untransformed H98 strain was grown on Hv-YPC media with and without thymidine, growth was only observed in media supplemented with thymidine. Transformation of the host strain was performed using both the unmethylated *fdx-phr2*-pTA233 and *fdx-phr1*-pTA233 constructs. A negative control (without any plasmids) with water was also included. The procedure was straightforward, although any rough handling of the spheroplasts would cause cell lysis so care was

taken when handling these cells. Also, if the cells lysed it would cause a PEG-NaCl precipitation of the DNA where transformation would not be possible (Cline et al. 1989). Transformed cells were streaked on both thymidine + and thymidine – plates and incubated after sealing using parafilm to avoid drying out (typically takes up to 6 days for colony formation). Visible red-orange colonies were seen just after 4 days on all the plates except the control where no colony growth was observed. On the thymidine + plates, the colonies were denser compared with the thymidine – plates. Single colonies were picked from thymidine – plates and inoculated into liquid Hv-YPC broth for 24-72 hr until the cells reach an OD_{600} of greater than 1.2. 1ml of the culture was stored as glycerol stocks in the -80°C freezer and the rest was spun down for purification of the plasmids. The *fdx-phr2*-pTA233 plasmids were subjected to double digests with *Nco*I and *Hin*dIII and electrophoresed on a 1% agarose gel (Figure 3.14). The *fdx-phr1*-pTA233 plasmids were digested with *Xba*I and *Hin*dIII and electrophoresed on a 1% agarose gel as well (Figure 3.15). Double digests shown in these figures indicate that transformations had been successful for both *fdx-phr2*-pTA233 and *fdx-phr1*-pTA233.



Figures 3.14: Gel electrophoresis analysis of double digest of H98 transformants containing *fdx-phr2*-pTA233 construct. From left to right: lane **1**&**11** = 1kb DNA ladder; **2**&**3** = clone 1, 10^0; **4**&**5**-clone 1, 10^1; **6**&**7** = clone 2, 10^0; **8**&**9** = clone 2, 10^1; **10** = undigested plasmid control; **12** = DNA mass ladder



Figures 3.15: Gel electrophoresis analysis of double digest of H98 transformants containing *fdx-phr1*-pTA233 construct. From left to right: lane 1&6 = 2 log kb ladder; 2-5 = digested plasmids

3.4.5 SDS-PAGE analysis of transformants

Both cell extracts and cell debris were prepared by sonication as per Section 3.3.7.1 using the *Hfx. volcanii* H98 strain transformed with pTA233 carrying the *phr* genes, H98 non-transformed strain, and the WT strain. The soluble (cell extracts) and insoluble (cell debris) fractions were analyzed by SDS-PAGE along with a sample of pure *E. coli* photolyase (Trevigen). There were no detectable differences in protein expression profiles between the transformed strain, untransformed strain and the wild type (Figure 3.16). The predicted M_r values for the putative Phr2 and Phr1 proteins are 54488.8 and 56511.8, respectively.

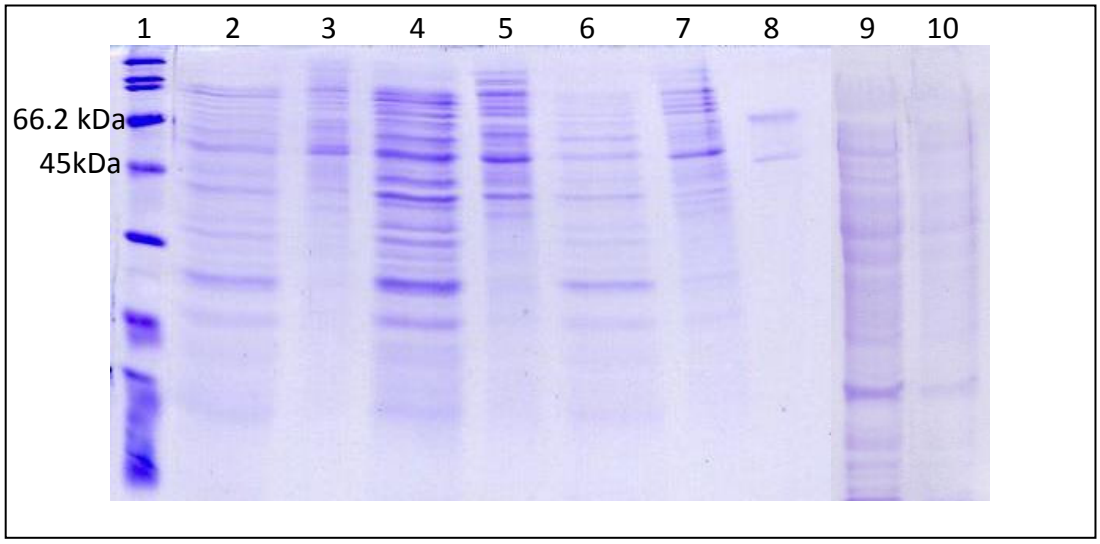


Figure 3.16: SDS-PAGE analysis of the *Hfx. volcanii* expressed Phr2 and Phr1 proteins. This SDS-PAGE showed cell extracts and debris of *Hfx. volcanii* H98 transformed with pTA233 carrying *phr2* (lanes 2 and 3) and *phr1* (lanes 9 and 10), untransformed H98 strain (lane 4 and 5), and WT strain (lanes 6 and 7). A pure *E. coli* photolyase (Trevigen) was also run as a benchmark (53669Da).

Lane	1	2	3	4	5	6	7	8	9	10
Fraction	m	s	i	s	i	s	i	EP	i	s

m = marker, s = soluble, m = insoluble, EP = *E. coli* photolyase

3.5 Discussion

Previous studies have highlighted the efficiency of the light repair system in halophilic archaea, including *Hfx. volcanii*, in combating the deleterious effects of UV on DNA. A study by McCready demonstrated that whilst the UV-induced dimers were repaired in the dark, white light illumination and repair of the damaged DNA was a much more efficient process in removing these lesions (McCready 1996). The genes homologues, *phr1* and *phr2*, were designated as putative photolyase-encoding genes from *Halobacterium* sp. NRC-1. McCready and Marcello reported the construction of a *phr2* deletion mutant which was UV-irradiated followed by subsequent exposure to visible light; results indicated the inability of the mutant strains to repair UV-induced damage, leading to a drastic decrease in its survival compared to the wild type strain (McCready and Marcello 2003). This led to the observation that the expression of the *phr2*-encoded protein was essential for photoreactivation to occur. A knock-out analysis of both *phr2* and *phr1* by Baliga and colleagues also showed the lethal effects of a *phr2* deletion, leading to the deduction that *phr2* encodes a functional photolyase (Baliga et al. 2004). No such

function, however, was observed for *phr1*. Nevertheless, the over-expression of the *phr*-encoded proteins for functional analysis has not yet been reported, most likely due to the technical difficulties in defining a compatible host for expressing halophilic proteins where the salt requirement is higher than the requirement of mesophilic proteins. In this chapter, the aim was to express the putative Phr proteins in a homologous system where the matching cytoplasmic conditions should contribute to the expression and production of soluble and active proteins for the functional characterization of both proteins.

The steady progression in archaeal molecular biology is facilitated by the development of tools for genetic analyses. It has been previously shown that archaea are insensitive to most antibiotics used in bacterial vector-host systems. It was later reported that mevinolin, an inhibitor of eukaryotic 3-hydroxy-3-methylglutaryl coenzyme A reductases, strongly inhibited this enzyme in haloarchaeal extracts, preventing cell growth in liquid media (Cabrera et al. 1986). A number of shuttle vectors have since been constructed since with the discovery of inhibition by mevinolin, followed by selection for nutritional requirement as well, for the transformation in both *Hfx. volcanii* and *E. coli*.

In this study, the *phr* genes passed through a number of cloning vectors before it reached the shuttle vector used to transform the native H98 strain. pGEM®-T Easy Vector System was used to enhance the copy numbers of the clones for sequencing purposes as well as to ensure that only fragments correctly digested with both restriction sites were used for subsequent ligation. The cloning procedure was complex due to the number of intermediate vectors used to reach the final transformation stage. The *phr1*-containing clones with the correct sequence ligated into the pIL 11 vector to recover the strong *fdx* promoter; the gene, along with the promoter were excised from the vector and inserted into the pTA233 shuttle vector. The *fdx* promoter of the *fdx* gene encoding the [2Fe-2S] ferredoxin in *Hbt. salinarum* is one of the strongest promoters available and shown to yield large activities even during the exponential growth phase in *Hfx. volcanii* transformants (Pfeifer et al. 1993). The cloning route for *phr2* was simplified as a construct previously used in the lab contained the same restriction sites for insertion into pTA233, therefore bypassing the pIL cloning step.

The *fdx-phr2*-pTA233 and *fdx-phr1*-pTA233 constructs were transformed into a *dam-/dcm- E. coli* strain before transformation into the H98 expression strain. Methylation is a modification system where the cleavage of DNA by endonucleases or restriction enzymes is blocked, thus preventing the destruction of cellular DNA. The *E. coli* chromosomes encode two types of methylases: dam and dcm methylases. Whereas the dam methylases recognize the DNA sequence GATC and methylates the adenine residue at the N^6 position (Marinus and Morris 1973), dcm methylases recognize DNA sequences CCA/TGG and add methyl groups to the C^5 position of the internal cytosine (May and Hattman 1975). *Hfx. volcanii* possess a restriction system that results in the degradation of these methylated adenine residues. It was previously shown that transforming methylated DNA into the native strain resulted in a 10^3 drop in transformation efficiency (Holmes et al. 1991). This obstacle may be overcome by passing shuttle vectors through a methyltransferase-deficient strain. Unmethylated plasmids, however, are maintained at a lower stability and so routine cloning should be carried out (Marinus 1987). The constructs should be transformed into a dam+ JM109 strain followed by passage through a methyltransferase deficient strain, JM110, just prior to *Hfx. volcanii* transformation.

The unmethylated constructs containing the both *phr2* and *phr*1 were transformed into *Hfx. volcanii* H98 strain, where the presence of the insert including the *fdx* promoter in pTA233 was confirmed by double digests of the purified plasmids. However, an over-expression of neither Phr2 or Phr1 protein was confirmed by SDS-PAGE, as there were no detectable differences in protein profiling between the soluble or insoluble fractions from the transformed H98 strain, the untransformed H98 (no pTA233 vector) and the wild type strains. These results can be explained by the low expression level of the recombinant protein, possibly due to the usage of the shuttle vector containing a low copy number. Presently to the author's knowledge, there are no published data on an efficient halophilic protein expression vector or on the characterization of a homologous halophilic protein expression system. A second possibility could be the degradation of the gene products post translation, though this is not expected as the proteins were expressed in their native environment.

Further purification, structural and functional analyses are complicated by the insufficient level of expression achieved for the Phr proteins. This suggests that the development for a high copy number vector in conjunction with with a tightly controlled inducible promoter for controlled expression, such as the tryptophan-induced (*tna*) promoter currently used to detect the expression of essential chaperonin protein in *Hfx. volcanii*, would be advantageous for the over-expression of halophilic proteins in a native host (Large et al. 2007). Combined with the favourable condition of high salt for maintaining protein structure and function, the homologous expression system would become a much more sensible choice for expressing halophilic proteins.

Purification of the Phr2 Protein

4.1 Introduction

Protein expression using the host organism is advantageous in terms of having a compatible cytoplasmic environment where the high salt is a requirement for the proper expression and function of proteins. In the previous chapter, homologous expression of *Hfx. volcanii* Phr2 protein was attempted using the native strain in the hope of obtaining a soluble, functional protein. Results indicated that it was difficult to detect an over expression of the Phr2 protein, an obstacle that was most likely due to the low copy number of the vector used during the cloning process. It would not be worthwhile to purify the protein while the expression level remained too low for detection on an SDS PAGE. To the author's knowledge, an efficient system of halophilic protein expression has not been described to date. The aim of this chapter, therefore, was to obtain a soluble protein through the affinity purification of a homologously expressed halophilic Phr2 protein from *Hfx. volcanii*. This involved the modification of the pTA233 shuttle vector to include a polyhistidine tag upstream of the cloned gene, followed by affinity purification of the tagged protein. This chapter describes both the methodology and logic behind the technique, followed by purification and mass spectrosocopy results.

4.2 Materials

4.2.1 Cell culture

Please refer to Section 3.2.1 for growth media.

4.2.2 Molecular biology studies

Oligonucleotide primers were obtained from MWG-Biotech AG, and eurofins MWG GmbH, Germany. Shrimp alkaline phosphatase was supplied by Fermentas Life Sciences. GoStar Master Mix was provided by GeneSys, Surrey, UK. The His•Bind Resin was purchased from Novagen, USA. Suppliers for all other materials can be found in Section 3.3.

4.3 Methods

4.3.1 Construct design

The construct was made so that an N-terminal His-tag preceded the *phr2* gene in a pTA233 vector. The *phr2*-pTA233 construct was made according to the cloning procedures outlined in Chapter 3, while the *phr2*-pET28a construct was made according to procedures from Chapter 5, both of which were as illustrated in Figures 4.1 and 4.2. The *phr2* fragment was digested from the *phr2*-pTA233 construct using *Nco*I and *Hind*III, where the pTA233 vector alone was purified and the *phr2* fragment was discarded. The *phr2*-pET28a construct, as shown in Figure 4.2, contained an *Nco*I restriction site upstream of the histidine sequence and the *phr2* gene; the construct was then digested with both *Nco*I and *Hind*III, where the *his-phr2* sequence was kept and purified. This fragment was ligated into the pTA233 shuttle vector to create a *his-phr2*-pTA233 construct containing the N-terminal His-tag as shown in Figure 4.3.

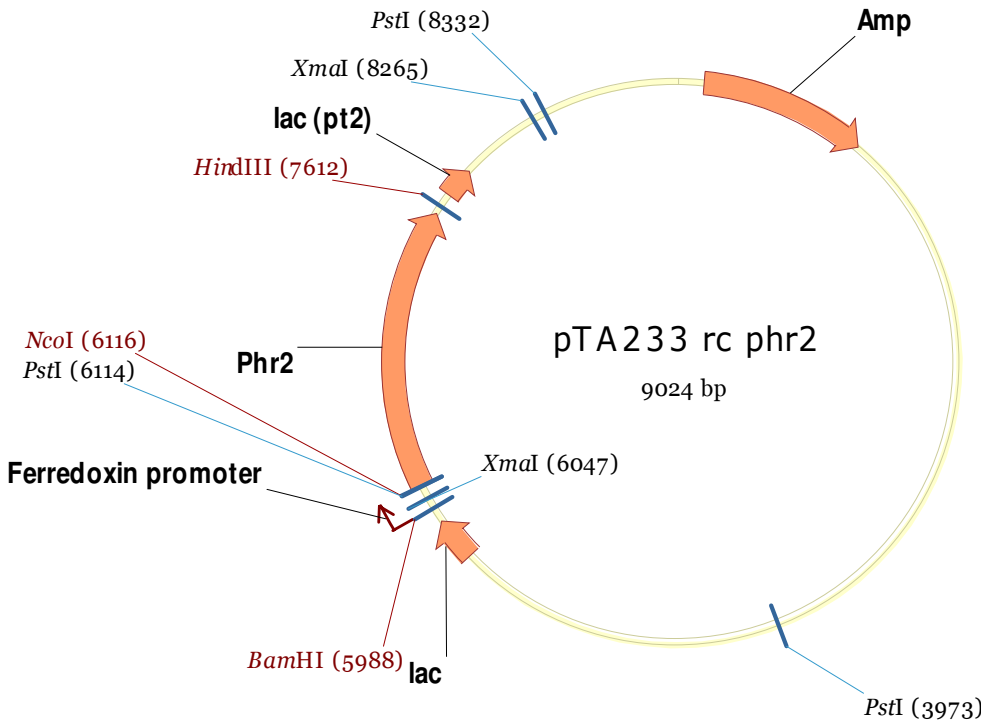


Figure 4.1: *fdx-phr2*-pTA233 construct. This fdx-*phr2*-pTA233 construct is as designed in Chapter 3. Restriction digests using *Nco*I and *Hind*III were used to remove the *phr2* gene only from the vector.

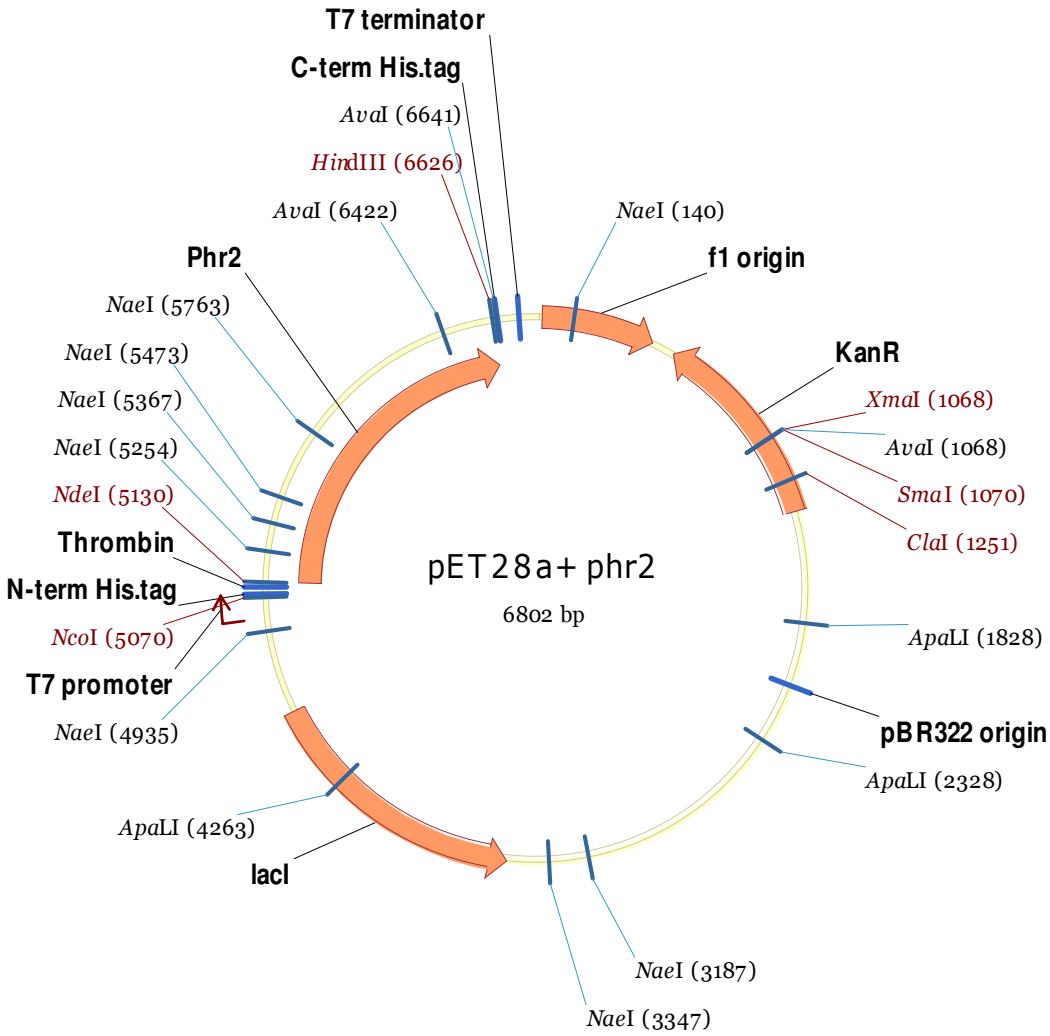


Figure 4.2: *phr2*-pET28a construct. This *phr2*-pET28a construct is as designed in Chapter 5. The sequence downstream of the *Nco*I site, including the His-tag (N-term His.tag) and upstream of the *Hind*III site was extracted from the vector.

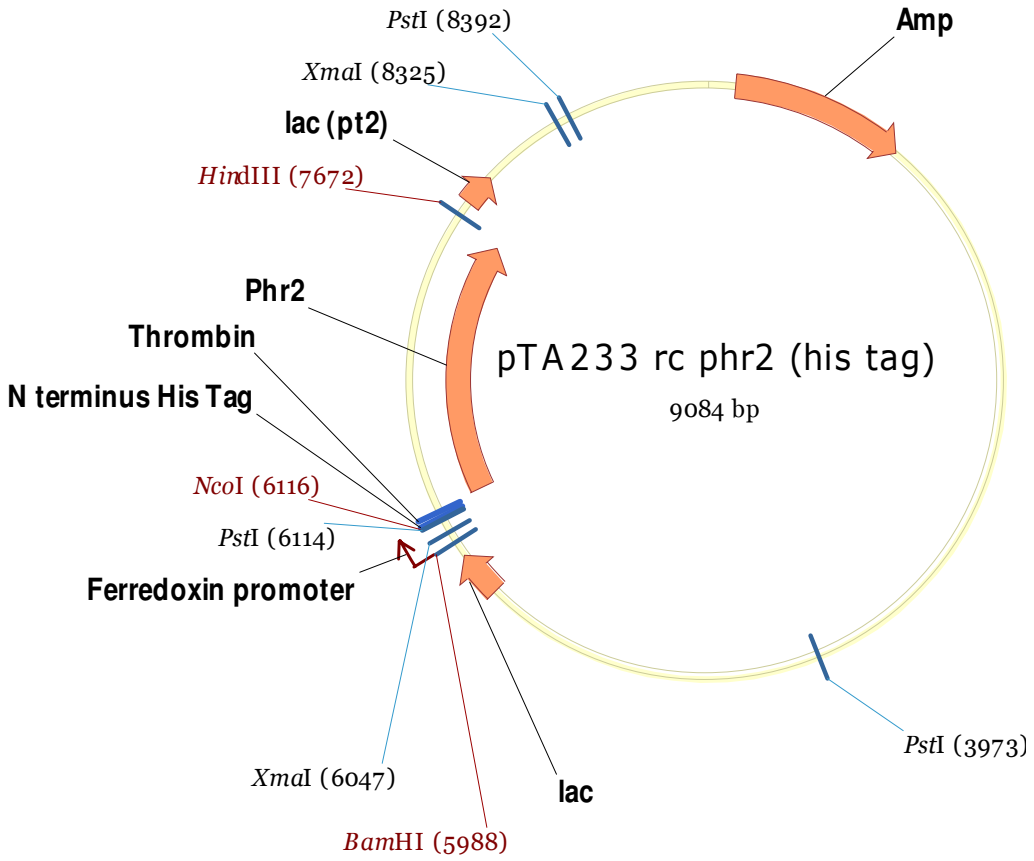


Figure 4.3: modified *his-phr2*-pTA233 construct. This is the proposed modification to the pTA233 construct. The *his-phr2* sequence digested from Figure 4.2 was cloned into the linearized pTA233 vector from Figure 4.1.

4.3.2 Transformation of *his-phr2*-pTA233 construct

The transformation protocol, plasmid prep purification, sequencing and single digest procedures can be found in Sections 3.3.2.6 to 3.3.2.9. The transformation of competent *dam-/dcm- E. coli* cells can be found in Section 3.3.4. The H98 strain transformation protocol was as followed in Section 3.3.6.

4.3.3 Colony PCR

Sterile toothpicks were used to gently swab the colonies grown on nutrient agar plates. The toothpicks were suspended into a 0.2 or 0.5µl PCR tube containing 50µl of sterile water to transfer the colonies into the tube. The tubes were heated on a heat block for 1 min, followed by centrifuging at 13,000 x g for 1 min. 1 to 2 µl was used as template for the PCR using the GoStar MasterMix.

4.3.4 Preparation of cell extracts for purification

10ml starter cultures of the transformed strains were set-up overnight in Hv-YPC medium, followed by inoculation into larger culture volumes (greater than 200ml) and grown for 2 days on shaking at 45°C. After two days, when OD_{600} reached beyond 1.2, the cultures were spun down at 8000 rpm for 15 min. The pellet was re-suspended in 5ml of extracting buffer (2M NaCl, 50mM Tris HCl pH8.0, and 1mM EDTA), followed by sonication of at least five 20s bursts (amplitude = 14 micron) in an ice bath with 40s of cooling between each burst depending on the viscosity of the re-suspended cultures. The cultures were sometimes diluted further in extracting buffer to loosen the pellet more and to make the sonication process easier. The sonicated cell extracts were centrifuged at 16,000 x g for 10 min to separate the soluble and insoluble fractions. The soluble fraction was filtered through a 0.45µm membrane to remove further contaminants.

4.3.5 Preparation of cell extracts for SDS-PAGE

Equal amounts of 2X SDS loading buffer was added to the filtered soluble fractions according to the OD_{600} readings taken earlier. The insoluble fraction was re-solubilized in 8M urea and spun down. 10 to 20µl of re-suspended fraction were mixed with an equal volume of 2X SDS loading buffer, again taking into account of the viscosity nature of the *Hfx. volcanii* cultures. The procedure for the analysis of transformants and the SDS-PAGE were as followed in Section 3.3.7.1 and 3.3.7.2.

4.3.6 Column preparation and purification

The column was prepared under renaturing conditions, i.e. without the addition of urea or guanidine HCl. The His•Bind Resin was gently mixed by inversion to re-suspend the slurry solution. 1ml of slurry was transferred to the gravity flow column. When the storage buffer dropped to the top of the column bed, the resin was washed 3 times with 5ml of sterile deionized water. The column was charged with 3 volumes of 8X (400mM) $NiSO_4$ solution until the resin turned a bright green. 3 volumes of binding buffer (50mM Tris-HCl, 2M KCl, 20mM imidazole, pH 8.0) was added and allowed to drain to the top of the column bed. 1 to 5ml of the prepared supernatant was loaded onto the column and allowed to drain through. The column was washed with 5 volumes of binding buffer, followed by eluting the bound

proteins with 5 volumes of elution buffer (50mM Tris-HCl, 2M KCl, 1M imidazole, pH 8.0). 1ml elution fractions were collected for each elution buffer used and were stored at 4°C for analysis. The column was washed with 5 volumes of sterile deionized water and stored in 20% ethanol.

4.4 Results

4.4.1 Construct design

4.4.1.1 PCR amplification of His-tag from pET vector

Different methods were sought for the modification to the pTA233 shuttle vector to include a His-tag. The first method involved PCR amplification of the His-tag sequence using the pET-28a vector as the template. The forward and reverse primers were designed to include *Nco*I restriction sites on both ends of the amplified fragment. The PCR would produce a fragment of approximately 50bp. As the *phr2* gene within the *fdx-phr2*-pTA233 construct was preceded by an *Nco*I restriction site, this meant that a single digest of the construct with *Nco*I would allow for the direct ligation of the amplified fragment with the same recognition sites on both N- and C-terminal ends. The primer sets used were listed in Table 4.1. The underlined sequence indicated the *Nco*I restriction site.

Table 4.1: Primers for insertion of the his tag sequence into *phr2*-pTA233

Primer	Sequence 5′ to 3′	Tm °C
pet28afwd	CCATGGGCAGCAGCCATCATCATCAT	**66.4**
pet28arev	CATGCCATGGGCTGCCGCGC	**67.6**
pet28a_for_prime	CACATGCCATGGCATCATCATCATCATCAC AGCAGCGGCCTG	**88.8**
pet28a_rev_prime	CACATGCCATGGGCTGCCGCGCGGCACCAG GCCGCTGCTGTG	**97.5**

Polymerase chain reaction was carried out as described in Section 3.3.2.2, where 5µl of each PCR reaction was run on a 1% agarose gel along with 1kb and 100bp DNA ladders as described in Section 3.3.2.3. Due to the small size of the PCR fragment generated, it was difficult to visualize the fragment on the agarose gel (Data not shown). A touchdown PCR was used to try to decrease non-specific background and to generate the desired fragment to a greater extent. However, due to the

unsuccessful visualization of the PCR products, it was not possible to carry out gel extraction of the desired fragment. 0.1% sodium tetrahydroborate buffer was used instead of TAE as a running buffer, as it has been shown in previous studies that the buffer helps with resolving small fragments on an agarose gel (Brody and Kern 2004). However, the 50bp fragment was still not observed; therefore, the PCR method was later discarded, and further design modifications were sought.

4.4.1.2 Designing synthetic oligonucleotides

A second method was utilised where synthetic oligonucleotides were designed for the direct insertion of the His-tag sequence into the *phr2*-pTA233 construct. This method involved designing oligonucletides containing both the His-tag and the thrombin cleavage site sequences with the appropriate restriction sites on both the N- and C-terminal ends. Again, *Nco*I was chosen as the recognition sequence directly preceding the start of the gene in the *phr2*-pTA233 construct. Two oligonucleotides were designed to be complementary, and contained overhangs to create "sticky ends" for the fragment to ligate into the linearized construct once it had been digested with *Nco*I. The first 4 oligonucleotides in Table 4.2a were used for ligating into the construct. The ratio of oligonucleotide to the linearlized *phr2*-pTA233 vector was optimized depending on the dilution factor used for the oligonucleotides; ratios of 3:1, 8:1 and 10:1 were chosen. Equal amounts of the complementary oligonucleotides (5µl of 100pmol/µl) were mixed in a 0.5µl PCR tube, and placed on a heat block at 95°C. To anneal the two oligonucleotides, the heat block was then turned off and allowed to cool to 60°C. Ligation reactions were set up to a final volume of 10µl. Transformation of the ligation reactions were done using JM109 competent cells, followed by plasmid prep and double digests using *Nco*I and *Hind*III. Colony PCR using primers hisphr2_fwd and hisphr2_rev (Table 4.2b) were performed to check for the success of the ligation by generating a 500bp PCR product including the his-tag sequence and part of the *phr2* gene sequence (Figures 4.4a and 4.4b). As only one restriction site was designed on either end of the oligonucleotide, re-ligation of the sequence was likely. To resolve this issue, the pet28a_for_phos and pet28a_rev_phos primers contained 5′ and 3′ phosphate modifications to prevent the re-ligation of the oligonucleotides. The cut pTA233 vector was also de-phosphoryated to prevent the re-ligation of the linearized construct after the single digest. PCR results from Figures 4.4a and 4.4b indicated

that PCR had been successful in producing the 500bp fragment. Colonies shown to produce the fragment of the right size were set up for overnight cultures, and plasmid DNA was purified and its sequence analyzed using the pTA233_forward primer (Table 4.2b). However, none of the plasmids contained the His-tag sequence; some contained parts of the *phr2* gene, while others contained random sequences that could not be aligned to the *phr2* sequence. After several failed sequencing results, this method was finally discarded.

Table 4.2a: Sequences of oligonucleotides and primers used

Oligonucleotides	Sequence 5′ to 3′
pET28-1_for	CCATGGCATCATCATCATCATCACAGCGGCCTGGTGCCGCGCG GCAGCCCATGG
pET28-1_rev	CCATGGGCTGCCGCGCGGCACCAGGCCGCTGTGATGATGATGA TGATGCCATGG
pET28-2_for	CATGGCATCATCATCATCATCACAGCGGCCTGGTGCCGCGCGG CAGCCCATG
pET28-2_rev	CATGGGCTGCCGCGCGGCACCAGGCCGCTGTGATGATGATGAT GATGCCATG
pet28a_for_phos	CATGGCATCATCATCATCATCACAGCAGCGGCCTGGTGC CGCGCGGCAGCC
pet28a_rev_phos	CATGGGCTGCCGCGCGGCACCAGGCCGCTGTGATGATGA TGATGATGC

Table 4.2b: Primers used for colony PCR and sequencing

Primer	Sequence 5′ to 3′	Tm °C
hisphr2_fwd	CATCATCATCATCACAGCGGCCTG	64.4
hisphr2_rev	ATGTCGAACTCTGCGTCGCCGTT	64.2
pTA233_forward	AATGCGGTTCGCGCGTTAATTGG	62.4

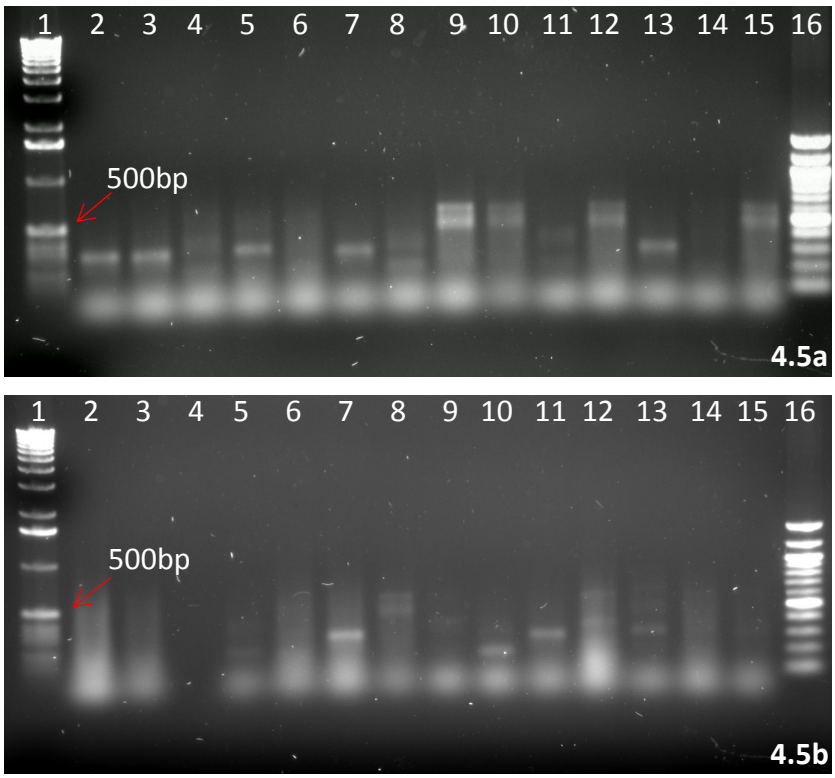


Figure 4.4a & 4.4b: Gel electrophoresis analysis of colony PCR of 30 possible transformants. **4.4a** – from left to right: lane **1**&**16** = 1kb and 100bp DNA ladders; 2-15 = plasmids 1-14. **4.4b** – from left to right: lane **1**&**16** = 1kb and 100bp ladder; 2-15: plasmids 15-28

4.4.1.3 Cloning of *his-phr2* fragment into pTA233 shuttle vector

The third method described here involved the cloning of the *phr2* gene into a pET28-a vector to produce an N-terminal his-tagged sequence, followed by a double digest of the His-tag along with the gene for insertion into the linearized pTA233 construct. This method, although more complex than the two previously described, produced the desired modification to the shuttle vector. The *phr2*-pTA233 construct was firstly subjected to a double digest with *Nco*I and *Hind*III to remove the *phr2* gene, whilst still maintaining the *fdx* promoter cloned from the pIL11 vector. A new PCR was initiated to amplify the *Hfx. volcanii phr2* gene for insertion into the pET28-a vector to produce an N-terminal his-tagged protein, where the forward and reverse primers contained restriction sites *Nde*I and *Hind*III, respectively. The complete *phr2*-pET28a cloning procedure can be found in the methods section in Chapter 5. Within the *phr2*-pET28a construct, the his-tag sequence was preceded directly by an

*Nco*I restriction site; a double digest was performed with *Nco*I and *Hind*III to obtain a "*his-phr2*" fragment, which was subsequently ligated into the linearized pTA233 shuttle vector, thus completing the modification scheme. Figure 4.5 shows the double digest results of both *phr2*-pTA233 and *phr2*-pET28a constructs using *Nco*I and *Hind*III. The linearized pTA233 vector (lane 6) and the *his-phr2* fragment (lane 3) were gel extracted and purified as shown in Figure 4.6.

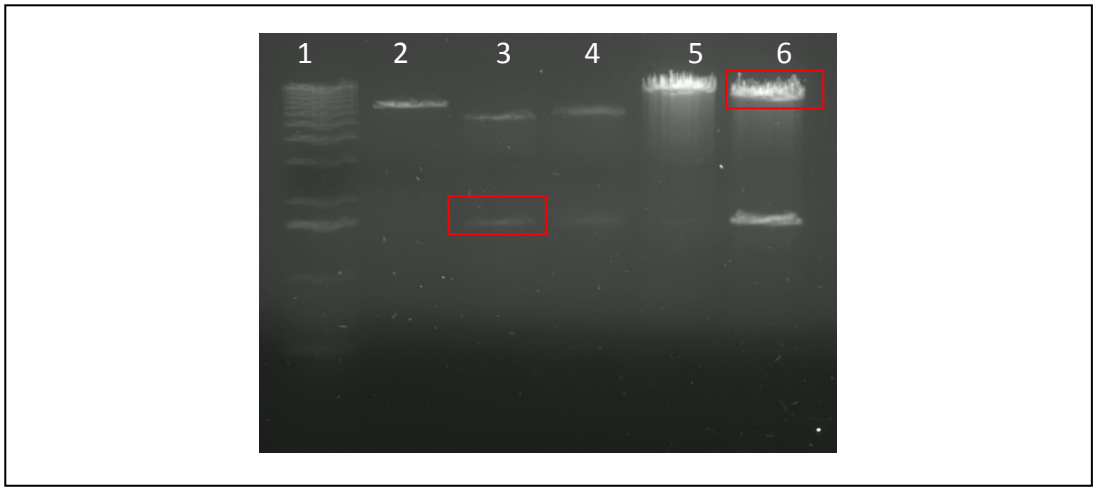


Figure 4.5: Gel electrophoresis analysis of single and double digest of *phr2*-pTA233 and *phr2*-pET28a constructs with *Nco*I, *Nde*I and *Hind*III. From left to right: lane **1** = 1kb DNA ladder; **2** = single digest (*Nco*I) *phr2*-pET28a; **3** = double digest (*Nco*I and *Hind*III) *phr2*-pET28a; **4** = double digest (*Nco*I and *Nde*I) *phr2*-pET28a; **5** = single digest (*Hind*III) *phr2*-pTA233; **6** = double digest (*Nco*I and *Hind*III) *phr2*-pTA233

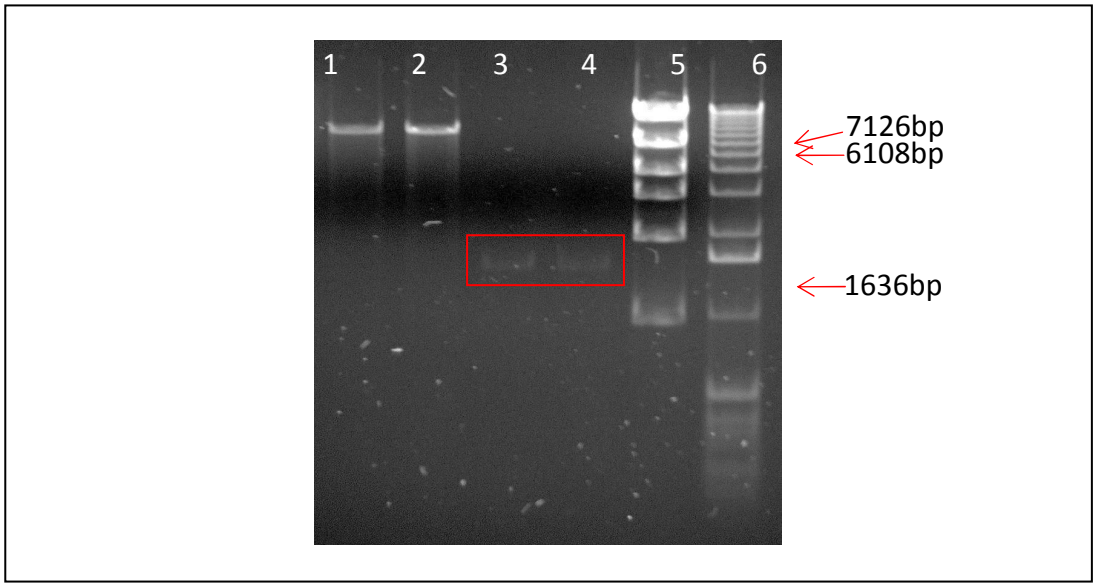


Figure 4.6: Gel extraction of the digested pTA233 vector and the digested *phr2* fragment from *phr2*-pET28a construct. From left to right: lane **1**&**2** = purified pTA233 vector; **3**&**4** = purified *his-phr2* fragment; **5** = DNA mass ladder; **6** = 1kb DNA ladder

Overnight ligation reactions were set up and transformation into highly competent JM109 cells was performed. 21 colonies were picked the next day for culturing over night, and all plasmids were purified using QIAprep Spin Miniprep Kit (Qiagen). A single digest with *Nde*I was used to check for successful ligations (Figures 4.7a and Figure 4.7b), as the newly modified *his-phr2*-pTA233 construct would contain the *Nde*I restriction site directly upstream of the *phr2* gene, whereas the *phr2*-pTA233 construct would not have contained this restriction site. As 17 out of 21 plasmids appeared to have been digested with *Nde*I, a representative 5 plasmids (1, 14 and 18 corresponding to lane 2 in Figure 4.7a and lanes 3 and 7 from Figure 4.7b) were chosen for sequencing. Only plasmids 14 and 18 showed the desired in-frame sequences containing the His-tag and the thrombin cleavage sequences upstream of the *phr2* gene. These two plasmids were used for subsequent transformation into competent *dam-/dcm- E. coli* JM110 cells as described in Section 3.3.4. The newly non-methylated *his-phr2*-pTA233 constructs were then transformed into H98 strain, protocol as followed in Section 3.3.6 Double digests with *Nco*I and *Hind*III were performed for the 5 plasmids transformed in H98 (Figure 4.8). Plasmid 1 (lane 2) was chosen due to its higher concentration shown on the 1% agarose gel compared to the rest of the plasmids, and was therefore used in the subsequent expression and purification experiments.

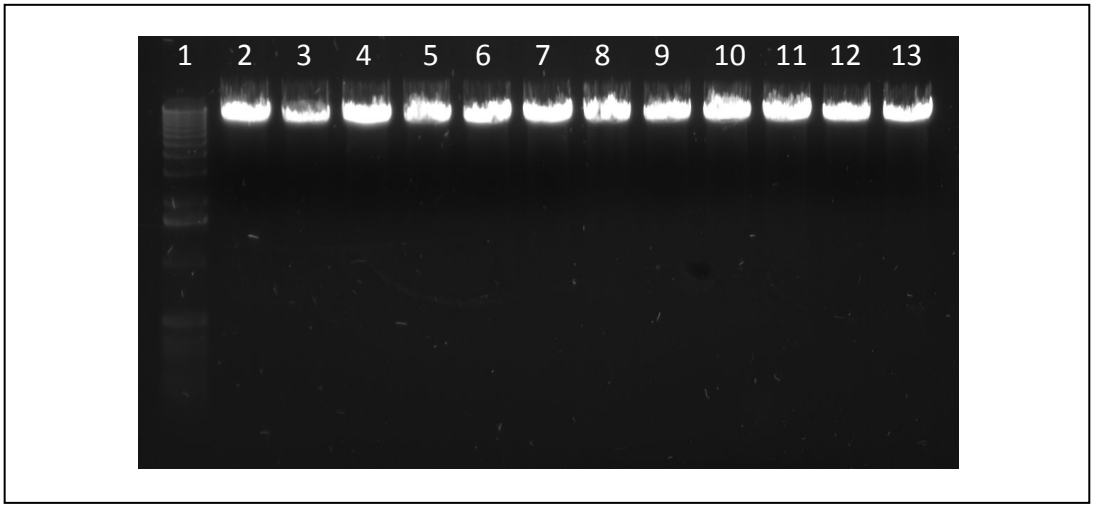


Figure 4.7a: Gel electrophoresis analysis of digested *his-phr2*-pTA233 using *Nde*I
From left to right: lane **1** = 1kb DNA ladder; **2**-**13** = plasmid preps 1 to 12

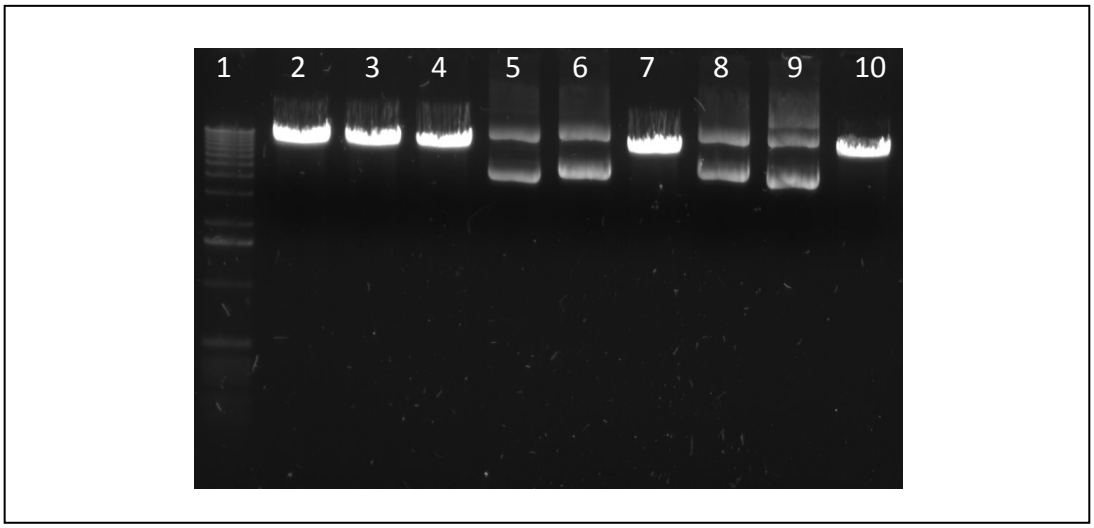


Figure 4.7b: Gel electrophoresis analysis of digested *his-phr2*-pTA233 using *Nde*I
From left to right: lane **1** = 1kb DNA ladder; **2**-**10** = plasmid preps 13-21

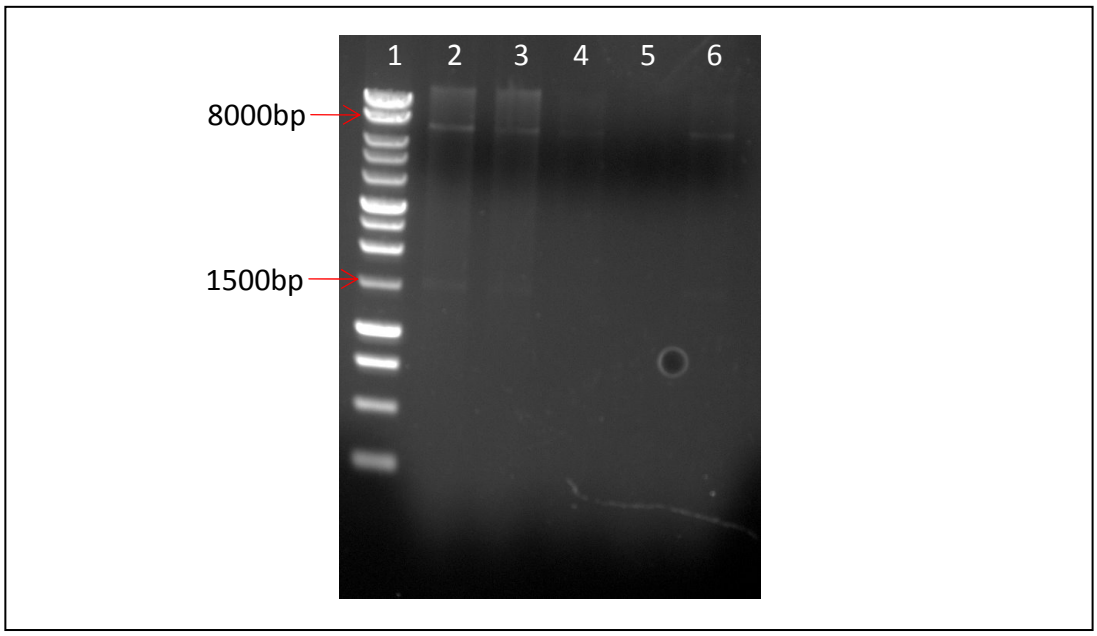


Figure 4.8: Gel electrophoresis of double digest of *Hfx. volcanii* H98 transformed *his-phr2*-pTA233 constructs. From left to right: lane **1** = kb ladder; **2**-**6** = plasmids 1 to 5

4.4.2 Protein expression and purification

Protein expression trials were performed for the homologously expressed his-tagged Phr2 protein, the homologously expressed non-tagged Phr2 protein, and a non-transformed H98 strain as a background control. 8µl of each sample was run on a 12.5% SDS-PAGE. Figure 4.9 shows that the expression profile did not differ much between the 3 strains, a result similar to that observed in Chapter 3. A pull down purification was then carried out for the His-tagged Phr2 protein using a His•Bind resin column charged with Nickel under renaturing conditions. To obtain enough proteins for the purification process, cells harvested from a 200ml culture was used with an OD_{600} of about 1.6. 1ml elution fractions were eluted and collected. Figure 4.10 shows the SDS-PAGE analysis of the purified fractions where 4µl of the eluted fractions were ran alongside the soluble and insoluble fractions on a 12.5% gel. Looking at this figure, there appeared to be a protein of approximately 56kDa in lanes 7 to 10, corresponding to fractions 6 to 9, with the most intense and cleanest band shown in fraction 7 (lane 8). This band was gel extracted and sent off for mass spectroscopy to confirm the identity of the purified protein.

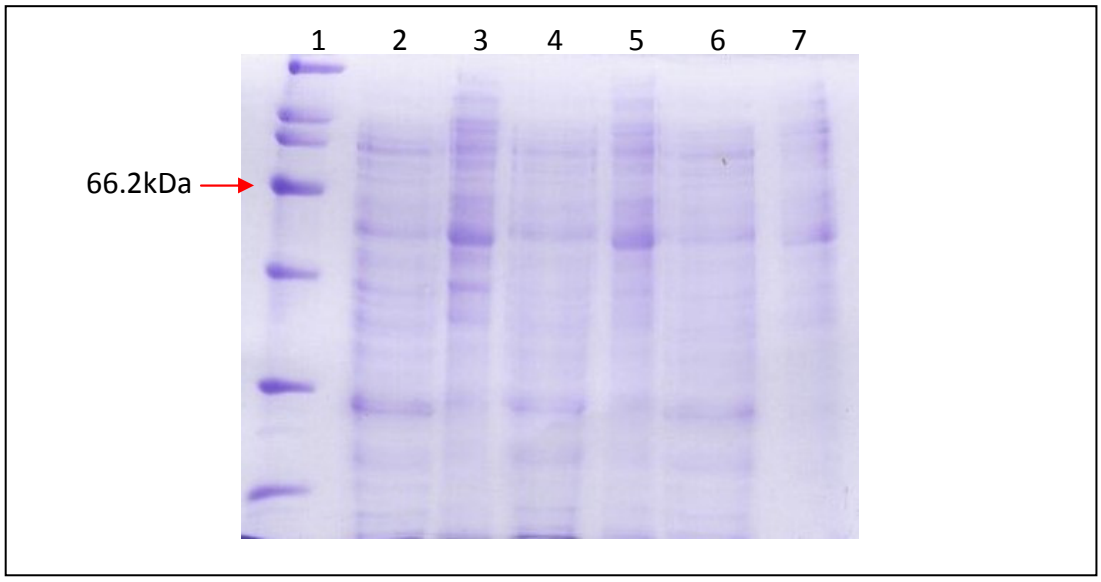


Figure 4.9: SDS-PAGE analysis of homologously expressed his tagged Phr2 protein, non-tagged Phr2 protein, and non-transformed H98 strain

Lane	1	2	3	4	5	6	7
Strain	-	T (tag)	T (tag)	T (no tag)	T (no tag)	NT	NT
Fraction	m	s	i	s	i	s	i

m = marker, **T(tag)** = his-tagged Phr2, **T (no tag)** = non-tagged Phr2, **Non-trans** = non-transformed H98, **s** = soluble, **i** = insoluble

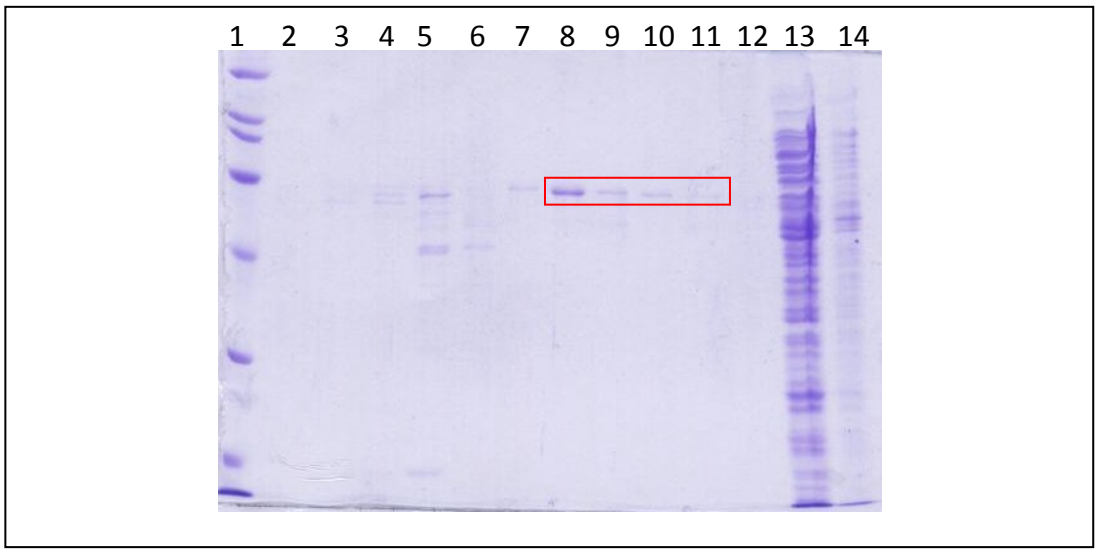


Figure 4.10: SDS-PAGE analysis of eluted fractions following purification of the homologously expressed His-tagged Phr2 Protein

Lane	1	2	3	4	5	6	7
Fraction	m	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}
Lane	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Fraction	7^{th}	8^{th}	9^{th}	10^{th}	11^{th}	i	s

m = marker**,** lanes **2-12 =** eluted fractions**, i =** insoluble**, s =** soluble

4.4.3 Mass spectroscopy of Phr2 protein

Due to the promising result shown by the SDS-PAGE on the purification of the Phr2 protein, the band of interest was excised from the gel and used for mass spectroscopy analysis along with the single amino acid sequence of the putative Phr2 protein. The analysis was conducted by the BMS Mass Spectroscopy and Proteomics Facility, University of St. Andrews, UK. The single protein was analysed by in-gel tryptic digestion and nanoLC-ESI MS/MS using the QStar Pulsar XL, a quadruple TOF instrument. The Mascot search result generated peptide identification in terms of probability, and this was compared against two databases: BMS (the facility's internal database) and the MSDB database for possible matches. Search against the BMS database did not result in a match to the Phr2 sequence, whereas a search against the MSDB database resulted in a match to three hypothetical proteins: RnnAC31100 from *Haloarcula marismortui*, NP22628 from *Natronomonas. pharaonis*, and VNG 2021 from *Halobacterium* sp. *NRC*-1. A BLAST analysis revealed that the three hypothetical proteins were homologues to the PitA protein from *Hfx. volcanii*. When the PitA protein was inputted into the BMS database, a significant match was produced between PitA and the analyzed sample. PitA has no sequence similarity to the Phr2 sequence.

4.5 Discussion

Soluble expression of the *Hfx. volcanii* Phr2 protein in the native strain was too low to enable purification (Chapter 3). The pTA233 shuttle vector used in the cloning process was not designed for expression, but rather as a selection marker system for the ability to utilize thymidine as a nutrient source (Allers et al. 2004). A vector designed specifically for the homologous expression of *Hfx. volcanii* proteins has not been described to date. An experimental approach was taken to identify a His-tag affinity purification method to purify halophilic proteins. While the method would not increase the expression levels of the protein, as an expression vector like pET28a would, it will provide the means to purify small quantities of Phr2 proteins expressed in the native strain.

Numerous strategies were utilized to modify the pTA233 vector, including PCR amplifying the His-tag sequence using the pET vector as the template. This method

was discarded as the amplified fragment was too small to be resolved on an agarose gel, even up to 2%. Alternative running buffers such as sodium tetrahydroborate, used to resolve small PCR fragments at a much higher voltage (over 200V), proved to be ineffective in resolving the 50bp fragment. A second method was conducted, where oligonucleotides of various lengths (56 to 61bp) containing the his-tag and the thrombin cleavage sequences were directly ligated into the linearized *phr2*-pTA233 fragment. Colony PCR using the transformed constructs were set up to determine whether the ligations had been successful. After numerous attempts, this method was also deemed ineffective, owning to the difficulties in ligating a small fragment into the vector.

A third method devised included further cloning steps and ligation of larger fragments into the vector. This would overcome both problems found in the previous methods: inability to visualize small fragments on an agarose gel and the low success rate of ligating small fragments into vectors. The *phr2* gene was successfully ligated into the pET28a expression vector as described in Chapter 5. The gene, along with the N-terminal his tag and thrombin cleavage sequences, was digested from the vector. At the same time, the *phr2*-pTA233 construct was digested to remove the *phr2* fragment, leaving the linearized pTA233 fragment for the *his-phr2* fragment to ligate into. This method proved to be a success; sequencing analysis indicated in-frame insertion of the his-tag, thrombin cleavage site as well as the *phr2* gene into the pTA233 vector. Purification results indicated that a protein of approximately 56kDa was eluted from the column, matching the size of the his-tagged protein as computed using the ProParam analysis tool from the Expasy Proteomics Server (http://www.expasy.ch/tools/protparam.html).

Mass spectroscopy analysis indicated that the eluted protein was not the expected Phr2, but a *Hfx. volcanii* protein, PitA. This encoded protein has a similar size of 56kDa, and possesses two domains: an N-terminal chlorite dismutase and an C-terminal antibiotic biosynthesis monooxygenase (ABM) domain, both on a single ORF as predicted by the Pfam database (Bateman et al. 2004). Three haloarchaeal proteins homologous to the PitA include the *Halobacterium* sp. NRC-1 Vng2021 (Ng et al. 2000), the *Haloarcula marismortui* RrnAC3100 (Baliga et al. 2004), and the *Natronomonas pharaonis* NP2262A (Falb et al. 2005). All 4 proteins possess a

similar domain architecture of two highly conserved regions, as well as a histidine-rich region predicted to be structurally unstable between the two regions (Bab-Dinitz et al. 2006). The purification of PitA instead of Phr2 could reflect different levels of expression of the two proteins, with higher levels of PitA expression compared to the His-tagged Phr2 protein. Whether PitA with its long string of histidine residues binds to the column at a higher affinity remains unknown. A further complication is the identical molecular weights of both proteins, suggesting that both proteins may have been co-purified, where PitA would out-compete Phr2 due to its higher expression level, or possibly, a stronger affinity to nickel binding. The predominant protein in this case was PitA as confirmed by mass spectroscopy analysis. The MSDB database also revealed that the peptide sequence was a significant match to the other three PitA homologues described.

While sequencing of the construct revealed that the *phr2* gene is correctly positioned with the His-tag, the low protein expression level of Phr2 remains an obstacle to the purification of this protein. This was complicated by the high expression of a native protein of a similar size and high binding affinity to the column. A future direction for continuing this experiment would be to utilize different tag systems. For instance, the use of a strep tag can potentially alleviate the problem of co-binding of another protein. Alternatively, a glutathione S-transferase (GST) fusion tag with a large size of 26kDa could potentially increase the solubility of the fusion partner, hence making purification more efficient (Waugh 2005). Making modifications to the homologous system, however, is a time-consuming procedure, where the use of different tags can also give rise to unforeseen problems downstream of either the cloning or the purification processes. For example, the insertion of larger tags, in contrast to the small protein peptide tags, may cause misfolding of the protein where there would be no binding site for the column-attached ligand. An alternative strategy must be sought for increasing the protein expression of Phr2. As discussed in the next chapter, a well-established and efficient system of expressing the target protein in an *E. coli* host was explored, followed by purification attempts of the heterologously expressed Phr2 protein.

Heterologous Expression in E. coli

5.1 Introduction

Heterologous expression in *E. coli* is a highly efficient method of obtaining high quantities of recombinant proteins, an advantage over the low level of protein expression achieved in a homologous system as demonstrated in Chapter 2. Firstly, the growth of *E. coli* is significantly faster than that of *Hfx. volcanii*, cutting down the time it takes to harvest the cells. Secondly, the understanding of the genetics and molecular biology associated with *E. coli* is more advanced than that of halophilic archaea. Thirdly, it has been shown that while expressing halophilic proteins in a low salt environment often results in soluble, inactive proteins or inclusion bodies that require reactivation and refolding, the protein yield is usually much higher than that of expression in the native host (Jolley et al. 1996; Connaris et al. 1999). This high yield would allow visualization of the over-expressed proteins, as well as the downstream purification and re-folding treatments as required. Also, the addition of a high-salt solution to an *E. coli* extract would be expected to precipitate many of the native *E. coli* proteins whilst leaving the halophilic proteins in a near-native environment. Studies with DHlipDH (dihydrolipoamide dehydrogenase) have shown that upon the addition of 2M NaCl, the heterologously expressed protein almost immediately restores its activity even after storage for 24 hr at 4°C in the absence of salt (Jolley et al. 1996). Other cases of the successful soluble expression of halophilic proteins using *E. coli* include the *Hbt. salinarum* nucleoside diphosphate kinase, which, interestingly, was found to maintain its activity and stability in the absence of salt even after 1 month storage at -20 °C (Ishibashi et al. 2001).

Soluble expression of the *Hfx. volcanii* photolyase genes proved difficult in native species (Chapter 4). This chapter therefore describes the results for expressing the Phr2 protein in a heterologous host, *E. coli*, in order to obtain high quantities of soluble, recombinant protein for the detection of enzymatic activity as well as attaining structural information on a halophilic photolyase. The *E. coli* pET expression system with the highly inducible promoter T7 was the chosen strategy for this study due to previous successes in expressing halophilic proteins, such as citrate

synthase and malate dehydrogenase, in soluble forms (Cendrin et al. 1993; Connaris et al. 1999).

5.2 Materials

5.2.1 Cell culture

Antibiotics kanamycin, chloramphenicol, streptomycin and gentamycin were obtained from Sigma Aldrich, Gillingham UK.

5.2.2 Molecular biology techniques

The pET28a vector, *E. coli* BL21 (DE3) and Rosetta™ expression strain was supplied by Novagen, USA. ArcticExpress™ was supplied by Stratagene. The Centricon concentrators were supplied by Millipore, UK. The suppliers for the rest of molecular biology materials were as mentioned in Sections 3.2.2 and 4.2.2.

5.3 Methods

5.3.1 Amplification of the *phr* genes

Oligonucleotide primers were designed to allow the incorporation of either N-terminal or C-terminal polyhistidine tag when inserted into a pET-28a vector. The method was as described in Section 3.3.1.1 which included the necessary checks for the compatibility of restriction sites used as well as for the melting temperatures of the primers. Forward primers were designed with either an *Nde*I or a *Nhe*I site as neither were present in the *phr2* gene sequence; this allowed the incorporation of an N-terminal His-tag following insertion into pET-28a using a reverse primer containing a *Hind*III site. To attach a C-terminal His-tag, the stop codon had to be removed from the gene and was replaced by random sequences containing a *Hind*III site for insertion into pET-28a. The forward primer contained an *Nco*I site to bypass the His-tag sequence directly downstream of this site in the vector map. Subsequently primers were also designed for the expression of the *phr1* gene product to check for its solubility in *E. coli*, as well as to serve as an expression control for *phr2*. All primer sequences can be found below in Table 5.1. The melting temperatures of any two sets of primers used in one reaction were within 5°C of each

other. The vector map and the cloning sequence of the pET-28a-c(+) expression vector can be found in Figure 5.1.

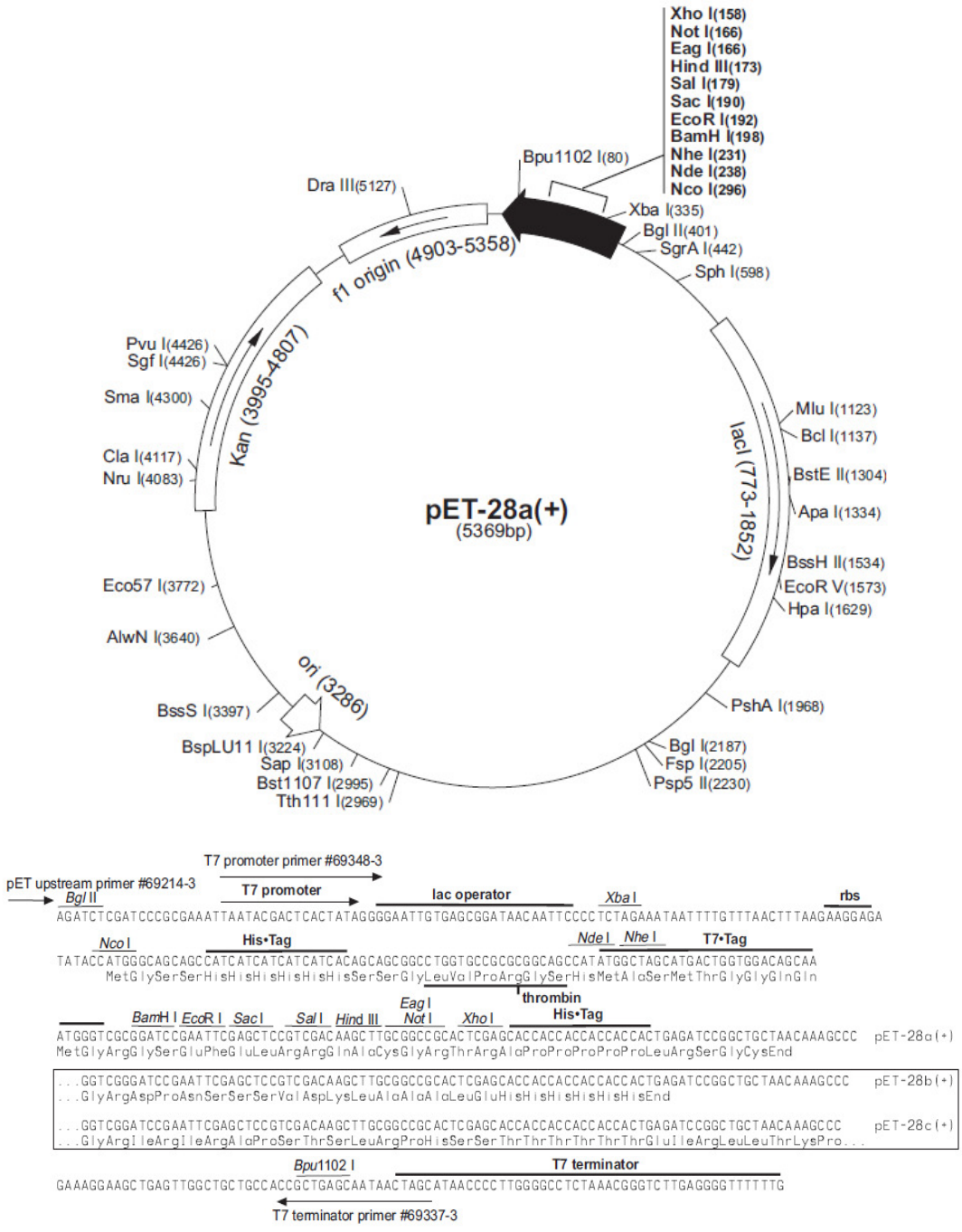


Figure 5.1: Vector map and the cloning/expression region of pET-28a-c(+)
On the map, the Kan^r denotes the kanamycin resistance gene. Ori is the origin of replication. The vector contains the multiple cloning sites for the insertion of the gene, a T7 promoter (370-386) upstream of the inserted clone between 198-238, a T7 transcription start at 369, a His•Tag coding sequence at 270-287, and a *Lac*I coding requence. The detailed view of the expression region was used to design primers to incorporate either an N- or C-terminal his tag.

Table 5.1: Primer sequences for amplifying both *phr2* and *phr1*

Primer	Sequence 5′ to 3′	Tm °C
HisF1-*Nde*I	GGAATTCGATTCCCATATGGAACTGCACTGG	68.2
HisF2-*Nde*I	CCCATATGGATTCGCCCATGGAACTGCAC	69.5
HisF-*Nhe*I	GGCTAGCCGATTCGCCCATGGAACTG	69.5
HisR-*Hind*III	AAGCTTAGTCGTCGCGTCGTCGCCGCG	72.6
Hfx F	GCGCCATGGAACTGCACTGGCACCG	74.9
Hfx R	CCAAGCTTGACCGAGGGCAAAAAACAGCAG	75.3
phr2rev*Hind*nonHis	AAAAGCTTAGTCGTCGCAAAGTCGTCGCC	69.7
phr2for*Nco*IHisC	TAATCCATGGAACTGCACTGGCACCGC	70.5
phr2rev*Hind*HisC	AAAAGCTTAGTCGTCGCTCAGTCGTCGCC	70.0
phr1for*Nde*IHisN	GGAATTCCATATGTCCGAGCCAACTTCG	67.5
phr1for*Nco*IHisC	TAATCCATGGCCGAGCCAACTTCGCTC	68.0
phr1rev*Hind*HisC	AAAAGCTTCGCGTGGGGCATATAATCGGA	66.1
phr1rev*Hind*nonHis	AAAAGCTTCGCGTGGGGCATCTAATCGGA	72.5

Polymerase chain reaction was carried out as described in Section 3.3.2.2 with the exception that different annealing temperatures were used based on the primer sets used. 5µl sample of each PCR reaction was run on a 1% agarose gel as described in Section 3.3.2.3. Amplified DNA fragments of the correct size were purified by DNA gel extraction using a QIAquick Gel Extraction Spin Kit (Qiagen). The purified PCR products were A-tailed using Taq polymerase followed by ligation into the pGEM®-T Easy vector. 1µl of the ligation reaction was used to transform 50µl of highly competent JM109 *E. coli* cells, where blue/white screening was used to select for transformants, followed by inoculation into 10ml LB media containing 50µg/ml carbenicillin. The cells were spun down and purified using the QIAprep Spin Miniprep Kit (Qiagen). Subsequent restriction digests were performed to screen for successful transformants, followed by sequencing analysis to check for errors within the sequences. Details for the subsequent experimental procedures (A-tailing, ligation, transformation, purification, sequencing and double digest) can all be found in Sections 3.3.2.4 to 3.3.2.9.

5.3.2 Preparation of competent JM109 *E. coli* cells

Frozen stock of JM109 *E. coli* cells was streaked unto an LB plate and incubated overnight at 37°C. A single colony was selected and inoculated into 10ml LB,

followed by incubation overnight at 37°C. The 10ml overnight culture was diluted into 200ml LB and grown to an OD_{600} of 0.6. The cells were pelleted at 4000rpm at 4°C for 10 min using 50ml Falcon tubes. The supernatant was decanted, and the cells were re-suspended in 200ml of ice cold 0.1M $MgCl_2$ by gently pipetting with a plugged pipette. The re-suspended cells were pelleted at 4000rpm at 4°C for 10 min, and the supernatant was decanted. The cells were re-suspended in 100ml of ice cold 0.1M $CaCl_2$ using 50ml Falcon tubes. The cells were placed on ice for 20 min, followed by pelleting at 4000 rpm at 4°C for 10 min. The cells were re-suspended in 10ml ice cold 0.1M $CaCl_2$ + 15% glycerol, and 100µl was aliquoted into pre-chilled sterile microfuge tubes on dry ice. The cells were stored at -80°C. The transformation efficiency was not checked as these cells were used for routine cloning purposes once the clones with the correct construct sequences were determined.

5.3.3 Protein expression

5.3.3.1 Preparation of cell extracts for time-course expression trials

10ml of transformed BL21 *E. coli* cell culture was incubated at 37°C overnight with shaking, then 0.5ml was taken from the culture and inoculated into 10ml LB media containing 30µg/ml of kanamycin the next morning. The number of cultures was determined by the number of time points and/or IPTG concentration used. The cells were grown to an OD_{600} of 0.4-0.6 and then induced at different IPTG concentrations (0.1mM, 0.5mM and 1mM) for mini expression trials. 1ml samples were taken at times 0, 2, 4, 6 and 24 hr. Samples were also taken from non-induced cultures as controls. The samples were spun down at 4000rpm for 5 min, supernatants were removed and the cell pellets were re-suspended in 100µl of extraction buffer (2M NaCl, 50mM Tris HCl pH8.0, and 1mM EDTA). Using the 150-W Ultrasonic Disintegrator (MSE Scientific Instruments), the extracts were subjected to three 15s bursts of sonication (amplitude = 14 micron) in an ice bath with 45s of cooling between each burst. The sonicated cell extracts were centrifuged at 16,000 x g for 10 min to separate the soluble and insoluble fractions. The supernatant (soluble fraction) was transferred to a clean 1.5ml microfuge tube; the cell pellet (insoluble fraction) was re-suspended in 100µl of sterile water. All samples were mixed with

an equal amount of 2X SDS loading buffer, and if needed, stored at -20°C for further analysis.

An alternative method of conducting expression trials was employed where a larger inoculation volume of 250ml was used prior to induction. This was to adjust for the low concentration of proteins observed when 1ml samples were taken. The culture was induced in the same way with the exception that a 10ml sample was taken at each time point. The 10ml sample was spun down as described, where the cell pellet was re-suspended in 1ml of extracting buffer before sonication.

5.3.3.2 Preparation of cell extracts for SDS-PAGE

10µl of samples (5µl fraction and 5µl loading buffer) were placed in the heat block for 5 min at 95°C and loaded on a 10% SDS-PAGE alongside BioRad Broad Range protein marker. The BioRad protein marker was prepared by diluting 1 in 10 with sterile water, followed by the addition of 2X SDS loading buffer in a 1:1 ratio and 2µl of 5% β mercaptoethanol. The mixture was boiled for 3 min at 95°C. The SDS-PAGE was run at 200V for 45 min. The gel was stained with Coomassie Blue, followed by de-staining overnight (Section 3.3.7.2).

5.3.3.3 Preparation of cell extracts for re-folding or purification

A 10ml starter culture, incubated at 37°C on shaking overnight, was added to 250ml LB media (not containing any antibiotics) and allowed to grow further until an OD_{600} of 0.6-0.8 (approximately 2-3 hr) before induction with 0.5mM IPTG . The culture was left in the shaking incubator overnight, and the cells were harvested by spinning down at 8000rpm for 20 min. The cell pellet was re-suspended in binding buffer (50mM Tris-HCl, 2M KCl, 20mM imidazole, pH 8.0) at a volume of 0.2g/ml, and sonicated 3 times at 20s burst in an ice bath with 40s cooling time between each burst. The sonicated fraction was spun down at 8000 rpm for 15 min to separate the soluble fraction from the pellet. The pellet was washed twice in extraction buffer and dissolved in solubilization buffer (50mM Tris-HCl, 2M KCl, 20mM imidazole, 8M urea, pH 8.0). The volume of solubilization buffer used depended on the viscosity of the pellet; generally the same re-suspension volume (0.2g/ml) was used

throughout the process. The pellet was mixed by vortexing followed by centrifuging at 16,000x g for 10 min. Further removal of insoluble material was done by filtering the supernatant through a 0.45 μm membrane prior to re-folding or purification. The resulting supernatant was stored at either 4 °C or -20°C for later use.

For preparing cell extracts to use on the AKTA™ chromatography system, the cell extract was concentrated using Centricon concentrators. The cell extract was poured into a concentrator and spun at 4000rpm for 10 min. The waste was poured off, and the procedure was repeated twice. 1ml of the final concentrated protein was used on the AKTA™ as per manufacturer's instructional manual. MOPS buffers were made up based on the estimated low pI of the Phr2 protein (Buffer A: 40mM MOPS, 2M NaCl and 20mM imidazole, pH 7.5. Buffer B: 40mM MOPS, 2M NaCl and 0.5M imidazole, pH7.5).

5.3.4 Purification

As purification was performed under denaturing conditions, all buffers used contained 8M urea. The His•Bind Resin was gently mixed by inversion to re-suspend the slurry solution. 1ml of slurry was transferred to the gravity flow column. When the storage buffer dropped to the top of the column bed, the resin was washed 3 times with 5ml of sterile deionized water. The column was charged with 3 volumes of 8X (400mM) $NiSO_4$ solution until the resin turned a bright green. 3 volumes of binding buffer (50mM Tris-HCl, 2M KCl, 20mM imidazole, 8M urea, pH 8.0) was added and allowed to drain to the top of the column bed. 1 to 5ml of the prepared supernatant was loaded unto the column and allowed to drain through. The column was washed with 5 volumes of binding buffer, followed by eluting the bound proteins with 5 volumes of elution buffer (50mM Tris-HCl, 2M KCl, 1M imidazole, 8M urea, pH 8.0). 1ml fractions were collected in microfuge tubes and stored at 4°C for further analysis. The column was washed with 5 volumes of sterile deionized water and stored in 20% ethanol. To re-generate the resin, however, the column was washed with 5 volumes of strip buffer (20mM Tris-HCl, 2M KCl, 100mM EDTA, pH 8.0), followed by washes with water and 20% ethanol.

5.3.5 Re-folding of purified protein

Various protocols were tested for maximizing the yield of re-folded proteins; step-wise dialysis was chosen for the least amount of protein loss due to aggregation. The purified fraction was first dialyzed in re-naturation buffer containing excess amount of FAD (20mM Tris-HCl, 2M KCl, 5mM $MgCl_2$, 2mM EDTA, 0.3mM GSSG, 3mM GSH, 1mM DTT and 4M urea) on a stirrer at 4°C overnight to allow for gradual buffer exchange and re-folding. The amount of FAD should be at least 10 times in excess to the amount of protein used. The following calculation, based on the Beer-Lambert Law, shows an example for calculating the amount of FAD needed.

For a 1ml Phr2 protein sample with a measured absorbance of 0.9:
$A = \varepsilon C L$ where A = Absorbance, ε = extinction coefficient, C = concentration, L = path length of the sample
$C = A / \varepsilon L$ where ε for Phr2 = $93405 M^{-1} cm^{-1}$ and length is 1cm
$C = 0.9 / (93405 M^{-1} cm^{-1}) (1cm) = 1.0 \times 10^{-5}$ M or 0.01mM
For FAD: 10 x 0.01 mM = **0.1mM** needed

The addition of 4M urea was to prevent the proteins from aggregating into precipitates prematurely. The next day, fresh buffer was used with 2M urea and no FAD, and the proteins were allowed to incubate overnight at 4°C on the stirrer. The final re-naturation buffer contained only 0.2M urea and 2M KCl, and the proteins were then again incubated overnight at 4°C, stirring. Protein concentration was determined by absorbance reading on the spectrophotometer at 280nm.

5.4 Results

5.4.1 Cloning of *phr* genes into pGEM®-T Easy Vector

5.4.4.1 PCR amplification of the *phr* genes

The *phr2* gene was amplified by PCR using various combinations of primers, which allowed the attachment of either N-terminal or C-terminal His-tags following expression in pET28a. The gene was successfully amplified from genomic DNA to produce fragments flanked by either *Nde*I or *Nhe*I and *Hind*III for an N-terminal His-tag attachment. Although a fragment flanked by *Nco*I and *Hind*III was produced for phr2 C-terminal His-tag attachment, subsequent sequencing revealed that there were base changes within the amplified gene, so it was not used for further experiments.

Figure 5.2 shows the PCR results obtained for producing *phr2* with an N-terminal His-tag attachment following expression. (Note: some of the initial PCR and expression trials were conducted by project student Richard Suckling).

Further PCR was performed for *phr2* and *phr1*, this time to produce the genes with C-terminal His-tag attachment (C-his *phr2*, C-his *phr1*) and no His-tag attachment (no-his *phr2*, no-his *phr2*), as well as an N-terminal His-tagged *phr1* (N-his *phr1*) (Figure 5.3). All PCR reactions were successful, as indicated by the amplified bands at ~1500bp for both *phr2* and *phr1* with gene lengths of 1455bp and 1512bp, respectively.

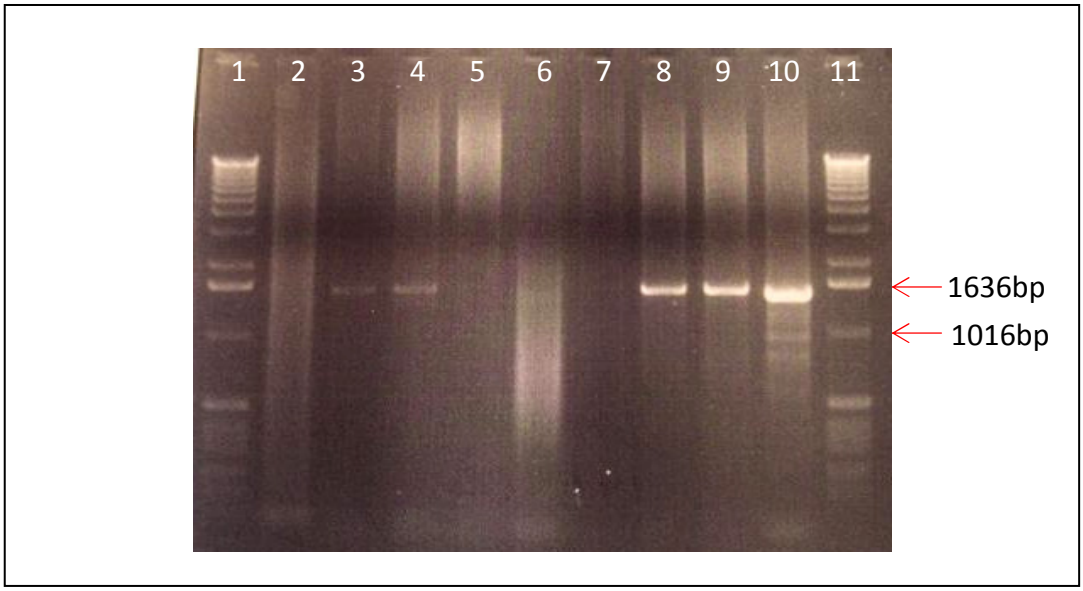


Figure 5.2: gel electrophoresis of PCR amplification of *phr2*. The *phr2* gene was amplified from both genomic DNA and a *phr2*-pGEM®-T construct. From left to right: lanes **1**&**11** = 1kb DNA ladder; **2**-**6** = amplified from genomic DNA; **7**-**10** = amplified from *phr2*-pGEM®-T construct.
Primers: **2** = negative control, **3** & **7** = HisF1–*Nde*I & HfxR, **4** & **8** = HisF2-*Nde*I & HfxR, **5** & **9** = HisF-*Nhe*I & HfxR, **6** & **10** = HfxF & HisR-*Hind*III.

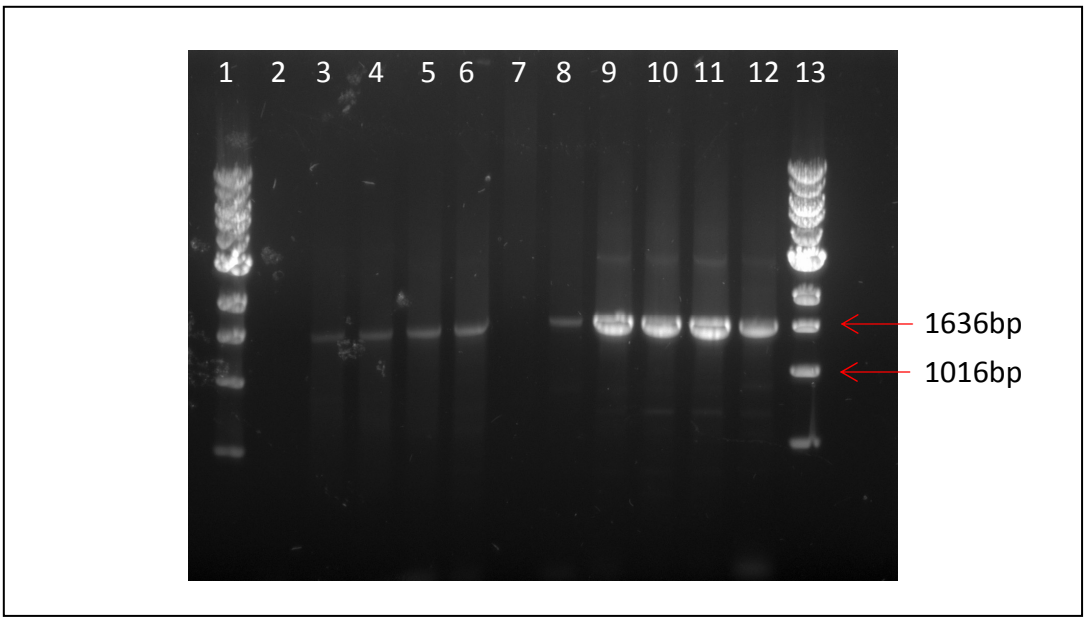


Figure 5.3: Gel electrophoresis analysis of PCR of *phr2* and *phr1*. From left to right: lanes **1**&**13** = 1kb DNA ladder, **2** = negative control, **3**&**4** = no-his *phr2*, **5**&**6** = C-his *phr2*, **7**&**8** = N-his *phr1*, **9**&**10** = C-his *phr1*, **11**&**12** = no-his *phr1*

5.4.4.2 Ligation into pGEM®-T Easy Vector

Following gel extraction of the bands, 6µl of each PCR product was modified through addition of adenine residues (A-tailed) to allow ligation into pGEM®-T. The ligation reaction products were transformed into commercial highly competent JM109 *E. coli* cells and plated on LB carbenicillin plates for blue/white screening. White colonies from each plate were picked and grown in liquid culture overnight.

5.4.4.3 Purification of plasmid DNA

All plasmids were purified using QIAprep Spin Miniprep Kit (Qiagen) and visualized on a 1% (w/v) agarose gel stained with ethidium bromide. Results are shown in Figures 5.4 and 5.5. For Figure 5.4, the plasmids in lanes 3, 4, 5, 7, 8, 10 and 11 appeared larger in size than those plasmids in lanes 2, 6 and 9, indicating possible successful insertion of *phr2* into pGEM®-T. In Figure 5.5, all colonies except for those pertaining to N-his *phr2* (colony 10, 11 and 12) appeared to have successful *phr2* and *phr1* insertions.

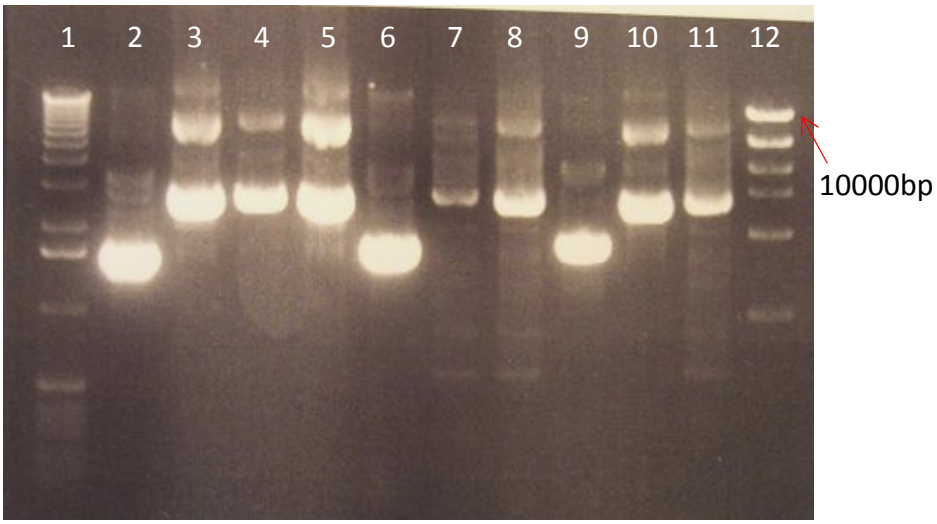


Figure 5.4: Gel electrophoresis analysis of purification of plasmids following ligation of *phr2* into pGEM®-T. From left to right: lane **1** = 1kb DNA ladder; **2** = HisF1-*Nde*I-G No.1; **3** = HisF1-*Nde*I-G No.2; **4** = HisF1-*Nde*I-G No.3; **5** = HisF1-*Nde*I-G No.4; **6** = HisF2-*Nde*I-G No.5; **7** = HisF2-*Nde*I No.6; **8** = HisF2-*Nde*I No.7; **9** = HisF2-*Nde*I No.8; **10** = HisF-*Nhe*I No.9; **11** = HisF-*Nhe*I No.10; **12** = High mass ladder

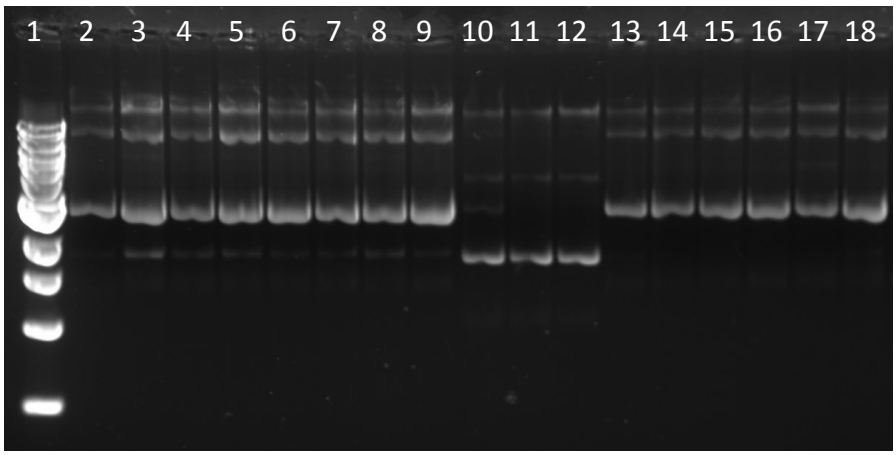


Figure 5.5: Gel electrophoresis analysis of purification of plasmids following ligation of *phr2* and *phr1* into pGEM®-T. From left to right: lane **1** = 1 kb DNA ladder; **2**-**6** = no-his *phr2*; **7**-**9** = C-his *phr2*; **10**-**12** = N-his *phr1*; **13**-**15** = C-his *phr1*; **16**-**18** = no-his *phr1*

5.4.4.4 Double Digest of plasmids

Restriction digests with *Nde*I and *Hind*III, as well as *Nhe*I and *Hind*III, were performed for N-his *phr2* prior to sequencing, results for which are shown in Figures 5.6 and 5.7.

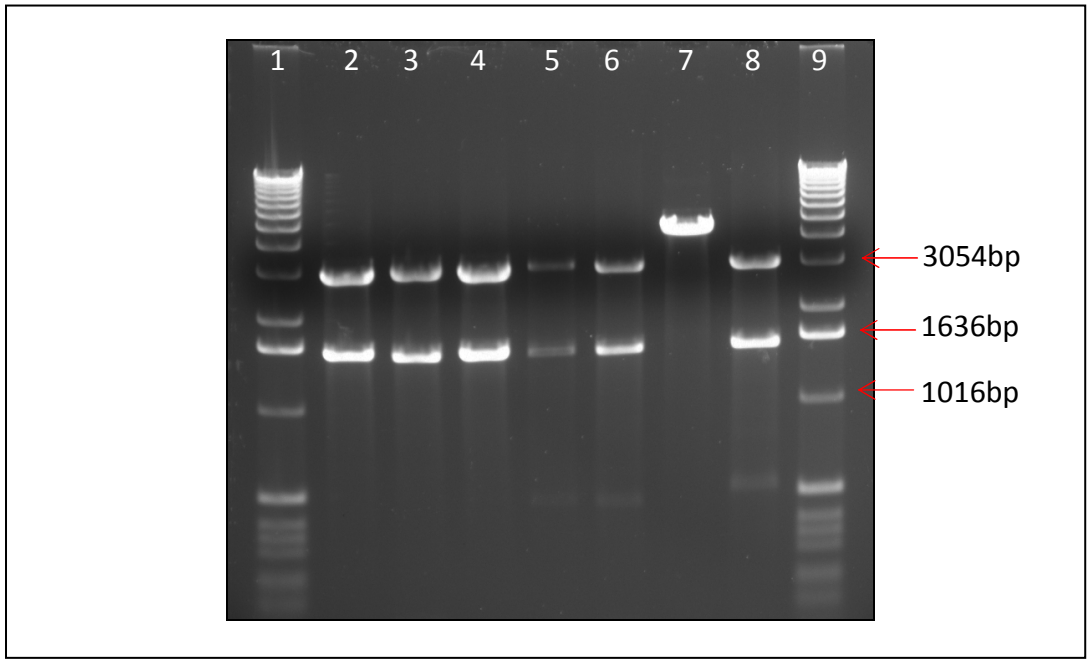


Figure 5.6: Gel electrophoresis analysis of the digested *phr2*- pGEM®-T constructs. From left to right: lanes **1**&**9** = 1kb DNA ladders; **2** = HisF1-*Nde*I-G No.2; **3** = HisF1-*Nde*I-G No.3; **4** = HisF1-*Nde*I-G No.4; **5** = HisF2-*Nde*I No.6; **6** = HisF2-*Nde*I No.7; **7** = HisF-*Nhe*I No.9; **8** = HisF-*Nhe*I No.10

In Figure 5.6, all of the plasmids except HisF-*Nhe*I No. 9 appeared to contain the *phr2* insert indicated by the two bands of ~1.5kb and ~3kb corresponding to the insert and the linearized pGEM®-T vector. The digest of HisF-*Nhe*I No. 9 produced a single band of ~4.5kb, indicating that the plasmid contains the insert but was digested by one enzyme only. However, as *Nhe*I and *Hind*III were efficient in producing two bands for the other reactions, mutation at one of the sites probably occurred. The sequences of the other 6 mini preps were analyzed.

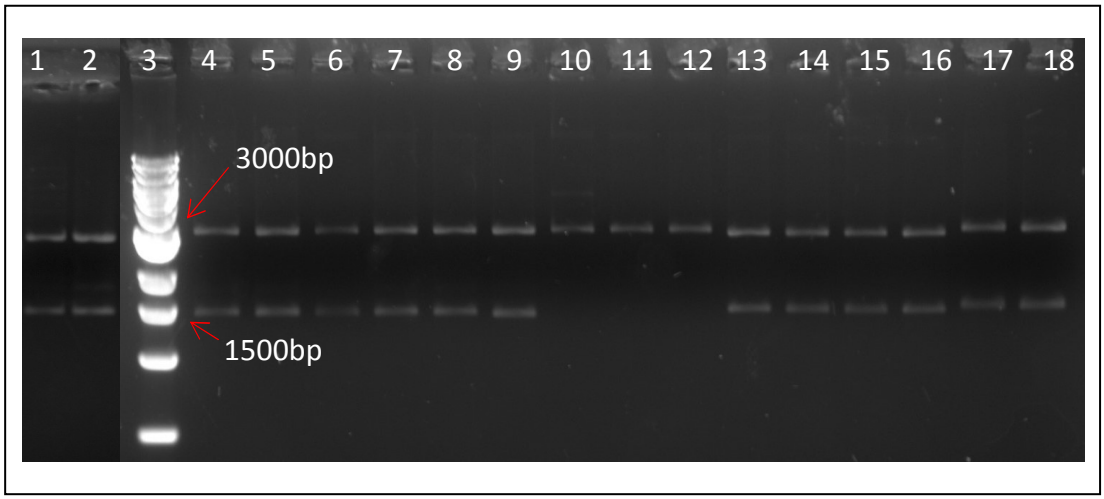


Figure 5.7: Gel electrophoresis analysis of the digested *phr2*-pGEM®-T and *phr1*-pGEM®-T construct. From left to right: lane **3** = 1kb DNA ladder; **1**-**2**, **4**, **5**-**6** = no-his *phr2*; **7-9** = C-his *phr2*; **10-12** = N-his *phr1*; **13-15** = C-his *phr1*; 16-19 = no-his *phr*

In Figure 5.7, double digests for plasmids in lanes 10, 11 and 12 corresponding to N-his *phr1* constructs were not successful as only one band at 3kb was produced. This indicated that ligation was unsuccessful for N-his *phr1*, and that the pGEM®-T vector was cut due to its internal restriction site for *Nde*I and re-ligation occurred as a result of the sticky ends produced by the single digest. A-tailing, ligation and transformation was repeated for this PCR reaction; however, ligations were again not successful. This led to new amplification reactions for *phr1* (Figure 5.8), and all bands corresponding to the right size (~1500bp) were excised and purified. A-tailing, ligation and transformation were repeated for PCR reactions in lanes 3 to 6 with no modifications to the protocols.

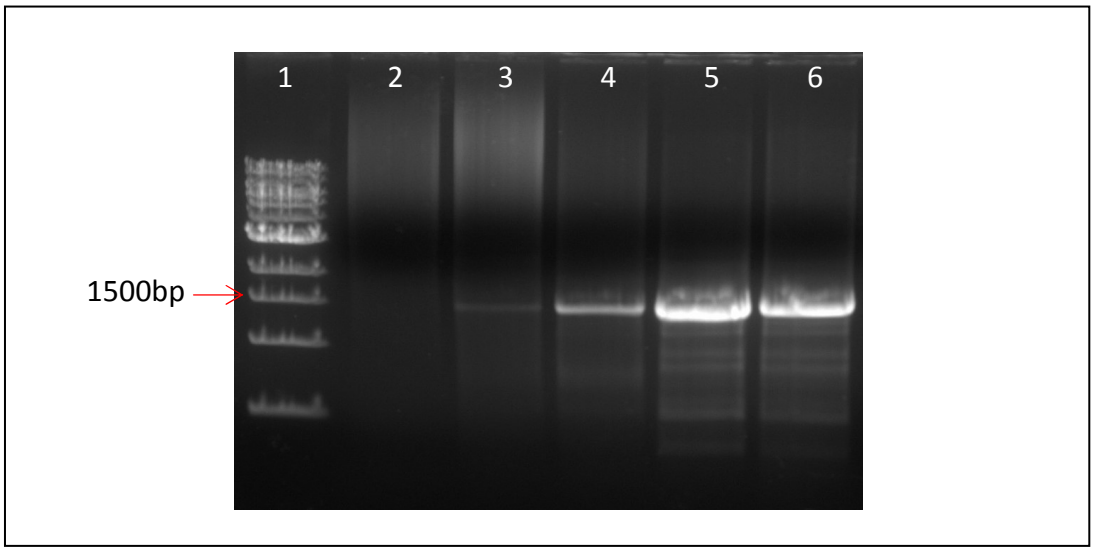


Figure 5.8: Gel electrophoresis analysis of new PCR for N-his *phr1*. From left to right: **1** = 1kb DNA marker; **2** = negative control; **3**&**4** = primers phr1forNdeIHisN and phr1HindRev; **5**&**6** = primers Phr1forNdeIHisN and phr1revHindnonHis

Post transformation, 8 white colonies were selected and liquid cultures were grown overnight. Plasmids were purified using QIAprep Spin Miniprep Kit (Qiagen). Double digest was performed using *Nde*I and *Hind*III and 5µl of the reactions were run and visualized on a 1% (w/v) agarose gel stained with ethidium bromide. In Figure 5.9, lanes 2 to 5 showed 3 bands at ~1500bp, ~3kb and ~4.5kb, corresponding to the *phr1* fragment, the pGEM®-T vector and the *phr1*-pGEM-T construct that was only cut with one of the enzymes, respectively. This was most likely due to the higher concentration of plasmids in these reactions where the amount of restriction enzymes added was not sufficient to digest all plasmids.

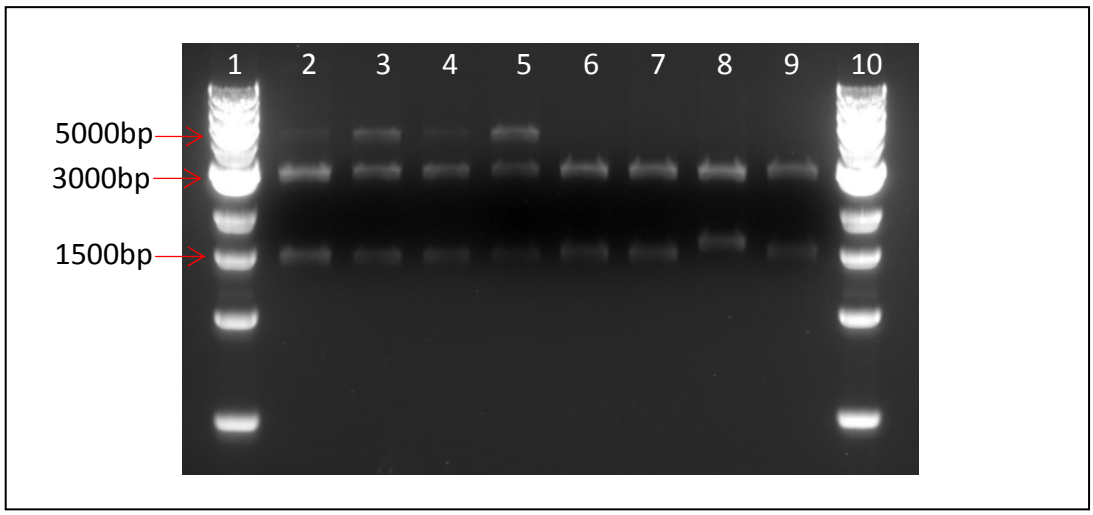


Figure 5.9: Gel electrophoresis analysis of the digested *phr1*-pGEM®-T constructs. From left to right: **1**&**10** = 1kb DNA ladder; **2**&**3** = PCR reaction 2; **4**&**5** = PCR reaction 3; **6**&**7** = PCR reaction 4; **8**&**9** = PCR reaction 5

5.4.4.5 Sequencing

The sequences of plasmids corresponding to lanes 2 to 6 and 8 from Figure 5.6 were analyzed using the SP6 and T7 primers for pGEM®-T, where results revealed that the HisF1-NdeI-G No.3 clone (lane3) was mutation free, apart from a base substitution within the primer sequence causing a substitution of the second amino acid from a glutamine to a glutamate. This mismatch occurred as the primer sequence was designed against the already made *phr2*-pGEM®-T construct and not against the *phr2* genomic sequence from *Hfx. volcanii*. This approach was taken as the gene was to be amplified from this construct initially; however, due to unsuccessful cloning it was decided to amplify the gene from genomic DNA. The forward primer originally had an *Nco*I restriction site and was replaced by *Nde*I for the purpose of attaching an N-terminal His-tag. This clone was chosen for subsequent cloning steps. Sequencing was performed for the other *phr1* and *phr2* constructs where the mutation free clones C1 (C-his *phr2*), C6 (no-his *phr2*), C14 (C-his *phr1*), C17 (no-his *phr1*) and Col 1 (N-his *phr1*) were chosen for ligation into pET28a vector.

5.4.2 Cloning of *phr* genes into pET-28a vector

5.4.2.1 Preparation of pET-28a *E. coli* expression vector

The pET-28a vector was transformed into commercial competent JM109 *E. coli* cells as per Section 4.3.2 and plated on LB kanamycin plates. Colonies were picked, inoculated into liquid cultures and grown overnight. The plasmids were then purified and 10µl of the purified samples of pET-28a plasmid was digested with *Nde*I and *Hind*III, *Nhe*I and *Hind*III, and *Nco*I and *Hind*III for ligation with either *phr2* or *phr1* inserts. Following digestion, the products were treated with shrimp alkaline phosphatase (SAP) to prevent re-ligation of the vector. The pET-28a digestion products were purified and 2µl of the eluted DNA was run on a 1% (w/v) agarose gel to quantify the amount of DNA and to ascertain that the double digests had been successful (gel electrophoresis not shown).

5.4.2.2 Preparation of *phr2* and *phr1* inserts

Restriction digests of the *phr2*-pGEM®-T and *phr1*-pGEM®-T constructs were repeated using 15 to 20µl of plasmid DNA to ensure ample amounts of the inserts was obtained following purification. The digested products were run on a 1% (w/v) agarose gel, followed by extraction and purification.

5.4.2.3 Ligation, purification and double digests

Ligation reactions were set up as described in Section 2.3.2.5, and incubated at 4°C overnight. The reactions were transformed into competent JM109 *E. coli* cells and plated on kanamycin plates for selection. Colonies were picked and grown in liquid culture overnight. The N-terminal His-tag *phr2*-pET-28a clones were then purified and visualized on a 1% agarose gel stained with ethidium bromide (Figure 5.10). The plasmids were subjected to double digests using the relevant restriction enzymes (Figure 5.11). Double digests revealed that all ligations had been successful as two bands of ~1.5kb and ~5.4kb was produced, corresponding to the insert and the linearized pET-28a vector, respectively.

The various C-terminal His-tag and non-tag versions of the *phr2* and *phr1* constructs were also ligated into pET-28a followed by transformation, purification of plasmids

and double digestion. As shown in Figure 5.12 and Figure 5.13, ligations had been successful for all constructs as two bands of the right sizes (insert at ~1500bp and the pET28-a vector at ~5000bp) were produced for the majority of the reactions.

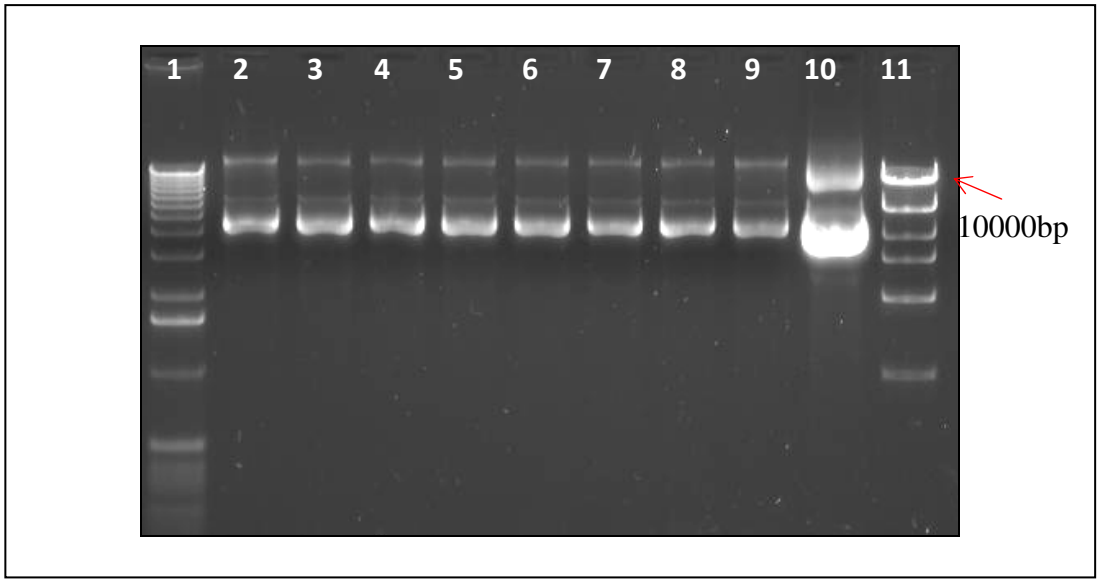


Figure 5.10: Gel electrophoresis analysis of purification of plasmid DNA following ligation of *phr2* inserts into pET-28a. From left to right: lane **1** = 1kb DNA ladder; **2** = HisF1-*Nde*I-G3 No.1; **3** = HisF1-*Nde*I-G3 No.2; **4** = HisF1-*Nde*I-G3 No.3; **5** = HisF2-*Nde*I7 No.1; **6** = HisF2-*Nde*I7 No.2; **7** = HisF2-*Nde*I7 No.3; **8** = HisF-*Nhe*I10 No.1; **9** = HisF-*Nhe*I10 No.2; **10** = HisF-*Nhe*I10 No.3

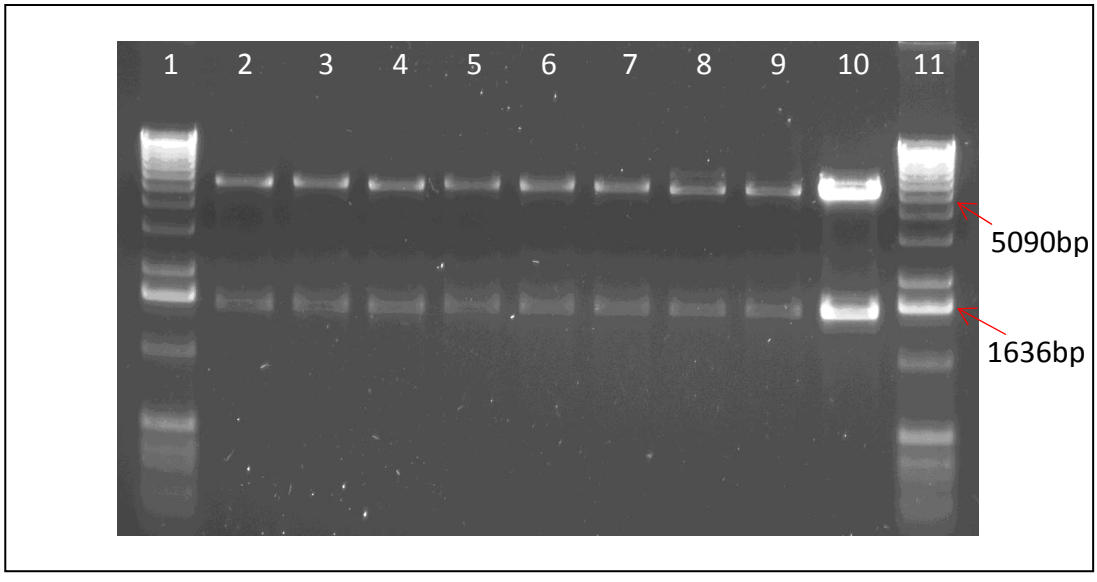


Figure 5.11 Gel electrophoresis analysis of the digested *phr2*-pET-28a constructs
From left to right: lane **1**&**11** = 1kb DNA ladder; **2** = HisF1-*Nde*I-G3 No.1; **3** = HisF1-*Nde*I-G3 No.2; **4** = HisF1-*Nde*I-G3 No.3; **5** = HisF2-*Nde*I7 No.1; **6** = HisF2-*Nde*I7 No.2; **7** = HisF2-*Nde*I7 No.3; **8** = HisF-*Nhe*I10 No.1; **9** = HisF-*Nhe*I10 No.2; **10** = HisF-*Nhe*I10 No. 3

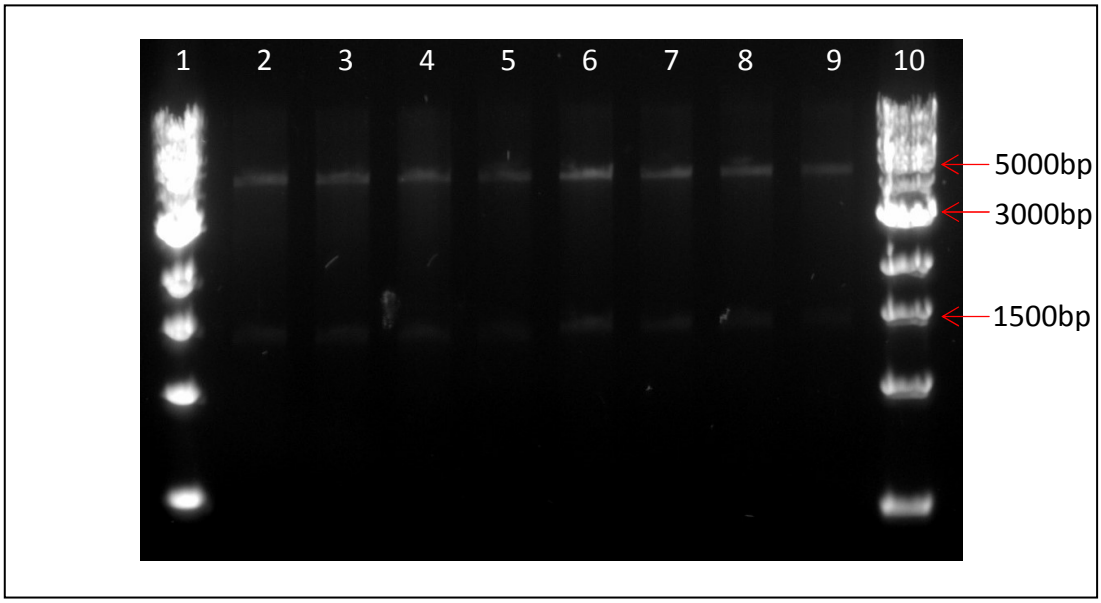


Figure 5.12: Gel electrophoresis analysis of the digested *phr2*-pET-28a and *phr1*-pET-28a constructs (C-terminal his tag and non-tagged constructs). Two colonies from each set of construct were selected and designated "a" or "b." The number referred to the plate number. From left to right: lanes **1**&**10** = 1kb DNA ladder; **2**&**3** = C-his *phr2*, 2 = C1a, 3 = C1b; **4**&**5** = no-his *phr2*, 4= C6a, 5 = C6b; **6**&**7** = C-his *phr1*, 6 = C14a, 7 = C14b; **8**&**9** = no-his *phr1*, 8 = C17a, 9 = C17b

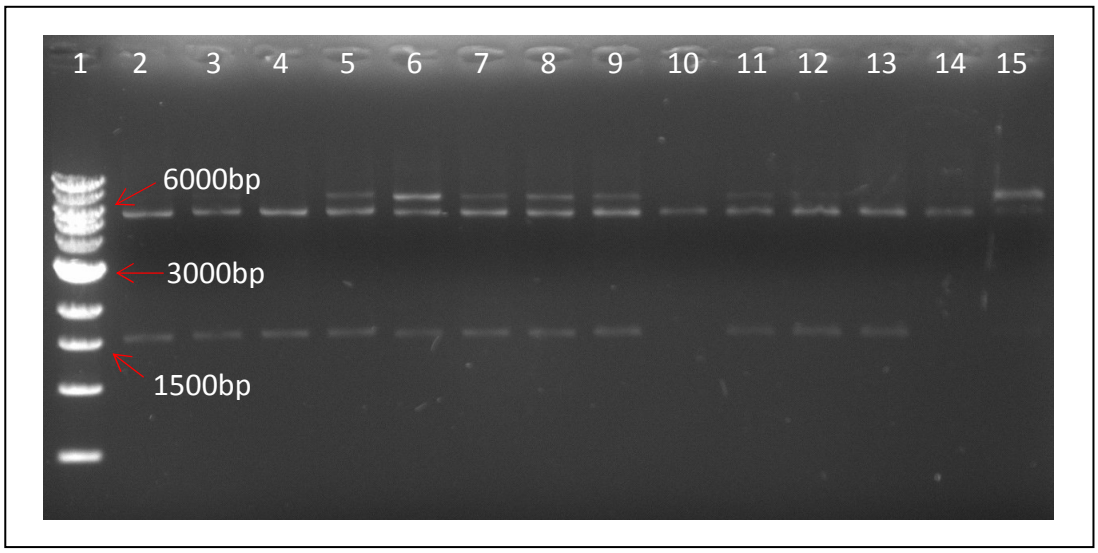


Figure 5.13: Gel electrophoresis analysis of the digested *phr1*-pET-28a constructs (N terminal his tag construct). From left to right: lane **1** = 1kb DNA ladder; **2**-**15** = 14 different plasmid preps used for restriction digest with *Nde*I and *Hind*III

5.4.2.4 Sequencing

Sequencing was performed for all constructs to ensure that the *phr* inserts were in-frame with either the N- or C-terminal His-tag sequence so to allow proper expression of both the gene and the tag. This was performed with T7 promoter, where the clones with the in-frame sequences were selected and used for protein expression trials.

5.4.3 Protein expression

Initial small-scale expression trials were conducted using the *phr2*-pET-28a construct. Following transformation of the construct into BL21 (DE3) cells, the soluble and the insoluble fractions were prepared post induction of protein expression by IPTG. Different IPTG concentrations (0.5mM and 1mM) were used to assess quantitatively the differences in recombinant protein expression using visualization on a SDS-PAGE. It was determined that at 1mM [IPTG], both the desired recombinant protein and the overall protein concentration were higher. Samples were also taken from non-induced cells as a control. The induction temperature was decreased from 37˚C to 30˚C to enhance the amount of soluble proteins produced, allowing more time for protein folding (Connaris et al. 1999). Samples were taken at times 0, 2, 4, 6 and 24 hr post-induction, and the cells were harvested for preparation of the soluble and insoluble fractions. Samples were run on a 12.5% SDS-PAGE gel, and proteins were stained with Coomassie Blue (Figure 5.14).

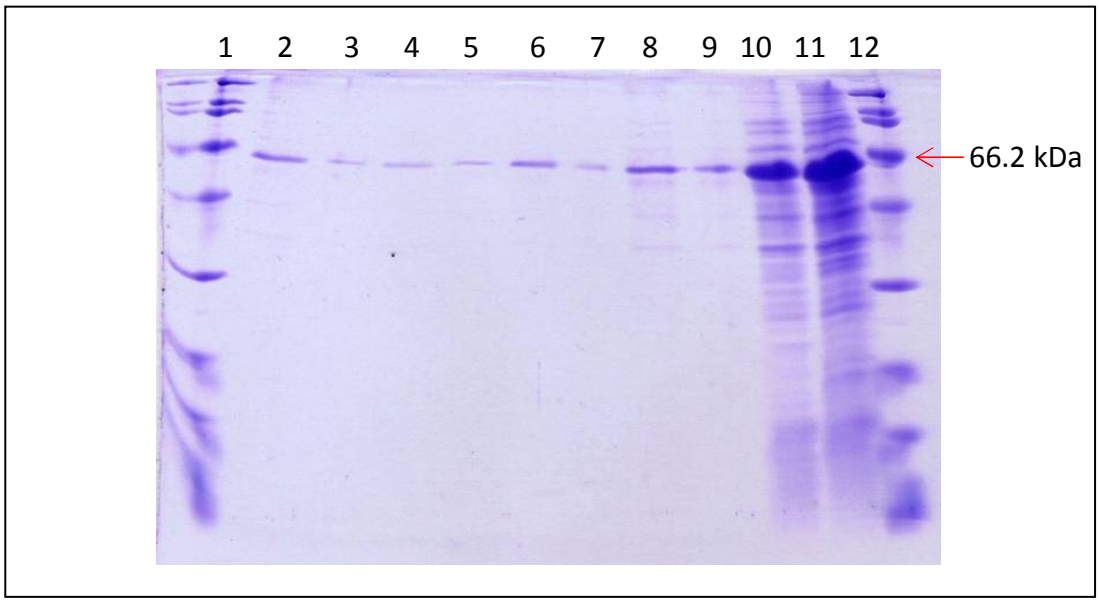


Figure 5.14: Protein expression analysis of N terminal His tagged Phr2 protein. 4µl of each sample (soluble and insoluble fractions) was run on a 12.5% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11	12
Fraction	m	s	i	s	i	s	i	s	i	s	i	m
Time (hrs)	-	0	0	2	2	4	4	6	6	24	24	-

m = marker; **s** = soluble fraction; **i** = insoluble fraction

In Figure 5.14, a band corresponded to a protein of ~60kDa was more intense than other bands at all time points in both the soluble and the insoluble fractions. The protein was estimated, using the Expasy ProtParam Tool (http://www.expasy.ch/tools/protparam.html), to have a molecular weight of 56318.7 Da, including the His-tag sequence. This band was most likely the recombinant over expressed His-tagged Phr2 protein. However, it did not appear that the Phr2 protein was more soluble after the decrease in temperature, as a large proportion of the protein was insoluble as seen, for example, in the 24 hr induced cultures. Nevertheless, there appeared to be sufficient amount of soluble Phr2 protein present in the sample taken at 24 hr to allow purification. A 250ml culture of transformed BL21 (DE3) cells was induced with 1mM IPTG and incubated at 30°C for 16 hr. The cells were spun down at 6000 rpm for 15 min and lysed in 10ml of extracting buffer. The soluble fraction was run through a His•Bind gravity flow resin column, charged with Nickel, for purification of the soluble His-tagged Phr2 protein. 1ml

fractions were eluted and collected in separate tubes. Figure 5.15 shows the SDS PAGE of the purification results obtained.

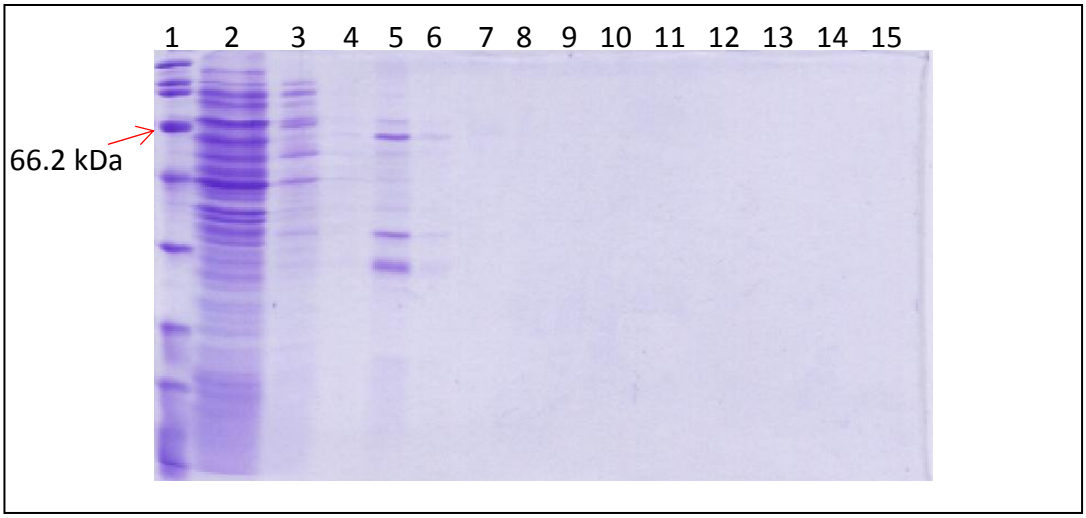


Figure 5.15: Analysis of eluted fractions following purification of over-expressed His-tagged Phr2 Protein. 4µl of each 1ml eluted fraction was run alongside the soluble cell extract, and the washed elutant, on a 12.5% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8
Fraction	m	Cell extract (soluble)	Wash elutant	1st	2nd	3rd	4th	5th
Lane	9	10	11	12	13	14	15	-
Fraction	6th	7th	8th	9th	10th	11th	12th	-

m = marker; lanes **4** to **15** = eluted 1ml fractions

From Figure 5.15, there appeared to be a protein of approximately 60kDa in lane 5 corresponding to the 2nd 1ml eluted fraction. However, the Phr2 protein concentration remained low as indicated by the lack of an over-expressed protein in the soluble cell extract seen in lane 2. To try to increase the quantity of protein, 10ml cultures of transformed BL21 (DE3) cells were induced with 1mM IPTG and incubated at 6 and 25 hr to replicate the previous over-expression. A 250ml culture using the Overnight Express™ Autoinduction system was also set up as a comparison. All fractions were run on a 12.5% SDS-PAGE as shown in Figure 5.16.

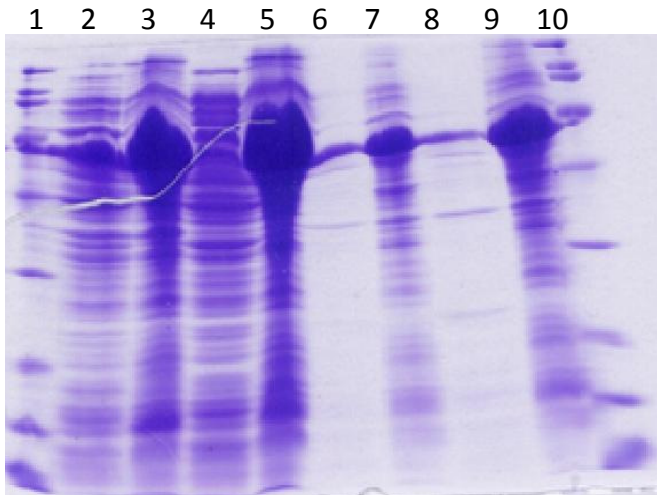


Figure 5.16: Protein expression analysis of His tagged Phr2 Protein. 4µl of each sample (soluble and insoluble fractions) was run on a 12.5% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10
Fraction	m	s	i	s	i	s	i	s	i	m
Culture volume (ml)	-	250	250	250	250	6	6	6	6	-
Media	-	LB	LB	OE	OE	LB	LB	LB	LB	-
Time (hrs)	-	16	16	16	16	6	6	24	24	-

m = marker; **s** = soluble; **I** = insoluble; **OE** = Overnight Express™ Autoinduction Media

As seen in Figure 5.16, the majority of the over-expressed Phr2 protein was insoluble in samples from all cultures. The over-expressed protein seemed more soluble at small volumes, as seen in the 6ml LB cultures induced at 6 and 24 hr. To purify the over-expressed protein at such small volumes would be a slow and inefficient process. Expression trials continued with varying growth temperatures, culture volumes and expression host strains. The AKTA™ purifier system was used to determine whether purification efficiency would increase. A 250ml culture of transformed BL21 (DE3) cells was incubated at 30°C overnight using the Overnight Express™ Autoinduction system. The cell extracts were prepared and a run was set up using a 1ml HiTrap™ chelating HP column. The fractions corresponding to the peak observed in the elution profile were run on a 10% SDS-PAGE. As shown in Figure 5.17, only a faint band was observed in the eluted fractions, and no protein over-expression was observed in the soluble cell extracts. It was therefore not

feasible to purify the protein on the AKTA™ until better soluble expression could be achieved.

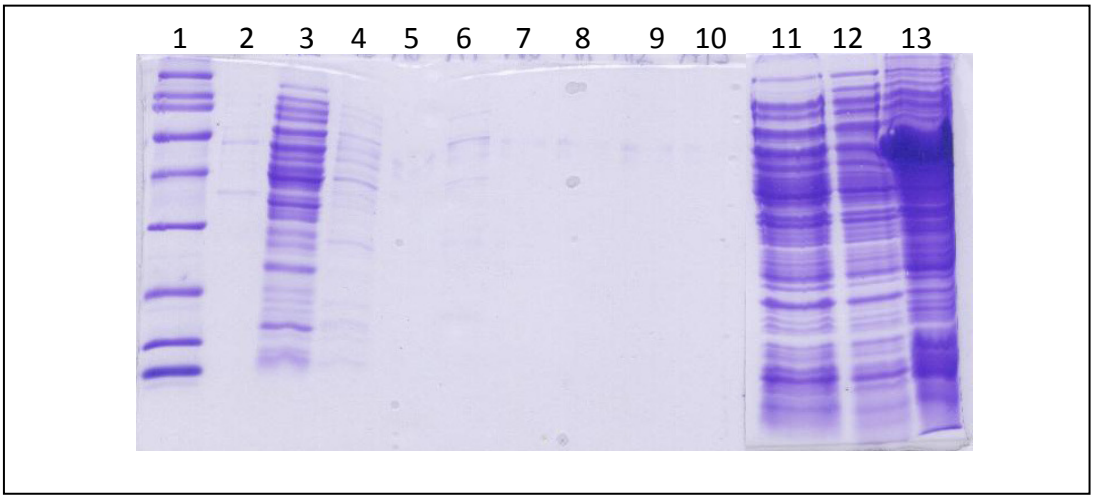


Figure 5.17: SDS-PAGE analysis of eluted fractions using the AKTA™ purifier system. 4µl of each 1ml eluted fraction was run alongside the soluble cell extract, and the washed elutant, on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11	12	13
Fraction	m	ft	ft	w	1st	2nd	3rd	4th	5th	6th	s	s	i

m = marker; **ft** = flow through; **w** = wash; lanes **5**-**10** = eluted fractions; **s** = soluble, **i** = insoluble

The induction trial was repeated using 25ml cultures of transformed BL21 (DE3) cells induced with 1mM IPTG incubated at 30°C, where 5ml sample volume was taken at each time point from both induced (Figure 5.18) and non-induced (Figure 5.19) cultures. Figure 5.18 clearly indicates that the protein was indeed expressed at 1mM IPTG but was largely insoluble even at 2 hr induction, and that basal induction levels were low as shown in Figure 5.19.

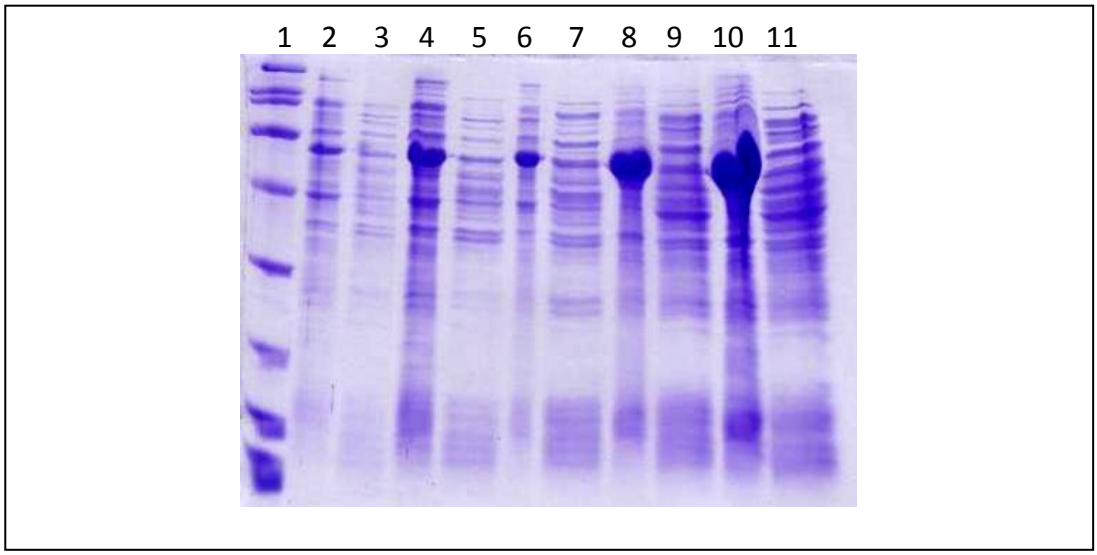


Figure 5.18: Protein expression analysis of N-terminal his tagged Phr2 protein. 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11
Fraction	m	i	s	i	s	i	s	i	s	i	s
Time (hrs)	-	0	0	2	2	4	4	6	6	24	24

m = marker; **i** = insoluble; **s** = soluble

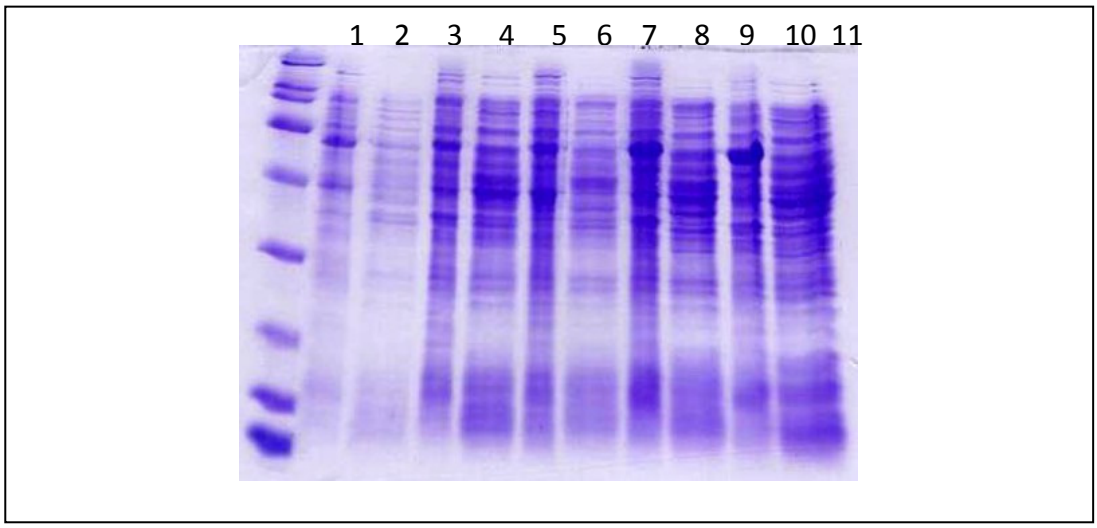


Figure 5.19: Protein expression analysis of N-terminal his tagged Phr2 protein. 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11
Fraction	m	i	s	i	s	i	s	i	s	i	s
Time (hrs)	-	0	0	2	2	4	4	6	6	24	24

m = marker, **i** = insoluble, **s** = soluble

Due to the formation of inclusion bodies using BL21(DE3), a different host strain Rosetta™, was trialled. The expression of proteins from thermophilic archaea, such as *Sulfolobus solfataricus* and *Thermoplasma acidophilum* can sometimes be difficult due to the differences in codon usage between *E. coli* and archaea (Wakagi et al. 1998; Kim and Lee 2006). Rosetta™, a BL21 host strain derivative designed to overcome the rare codon usage problem (Novy et al. 2001), was used to conduct expression trials at two different IPTG concentrations (1mM and 0.5mM) at 15°C. The original Rosetta strains supply tRNAs for the codons AGG and AGA (arginine), AUA (isoleucine), CUA (leucine), CCC (proline), and GGA (glycine) on a compatible chloramphenicol-resistant plasmid, pRARE. A scan of the *Hfx. volcanii* Phr2 nucleotide sequence indicated that whilst only a few rare codons were present and no tandem repeats of arginine were found, proline coded by CCC was abundant. The N-terminal His-tagged *phr2*-pET28a construct was transformed into the Rosetta™ strain and the colonies were selected based on their antibiotic resistance to kanamycin and chloramphenicol (Figures 5.20 and 5.21). The conditions for expression in the Rosetta™ strain were the same as for BL21 (DE3), found in Sections 5.3.3.

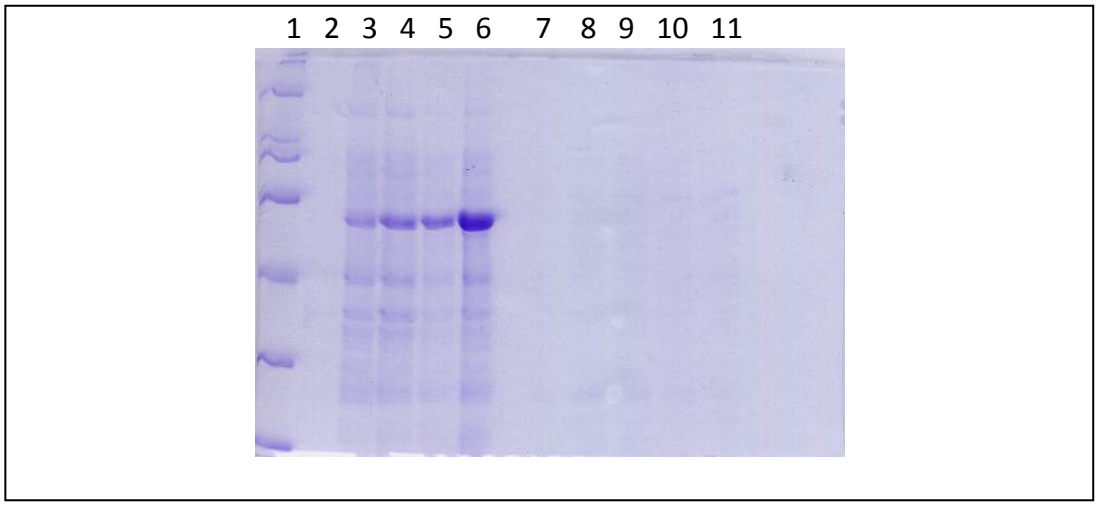


Figure 5.20: Protein expression analysis of his tagged Phr2 protein using Rosetta™. Transformed Rosetta™ cells were induced with 1mM IPTG and incubated at 15°C. 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11
Fraction	m	i	i	i	i	i	s	s	s	s	s
Time (hrs)	-	0	2	4	6	24	0	2	4	6	24

m = marker, **i** = insoluble, **s** = soluble

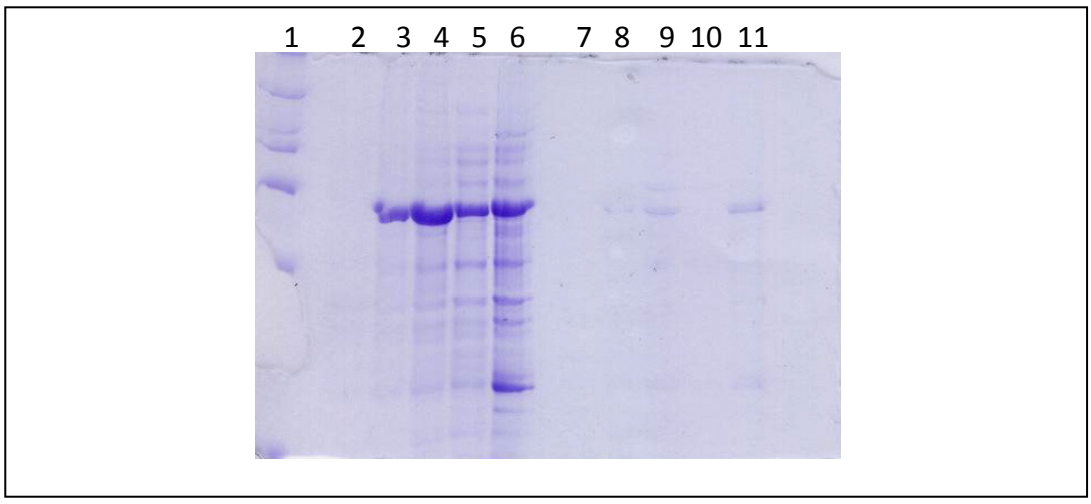


Figure 5.21: Protein expression analysis of his tagged Phr2 protein using Rosetta™. Transformed Rosetta™ cells were induced with 0.5mM IPTG and incubated at 15°C. 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11
Fraction	m	i	i	i	i	i	s	s	s	s	s
Time (hrs)	-	0	2	4	6	24	0	2	4	6	24

m = marker; **i** = insoluble; **s** = soluble

Figures 5.20 and 5.21 demonstrate the insolubility of the His-tagged Phr2 protein using Rosetta cells. However, reducing the temperature to 15°C slightly enhanced solubility at 0.5mM IPTG as shown in lane 11 from Figure 5.21. Therefore, a new strain, ArcticExpress™, was trialled for expressing the Phr2 protein at 12°C. These *E. coli* cells contain cold-adapted chaperonins, cpn60 and cpn10, derived from an Antarctic isolate, *Oleispira Antarctica*, known to facilitate proper protein folding by binding to and stabilizing unfolded or partially fold proteins. Cpn60 and cpn10 enhance the cell's ability to grow at low temperatures and allow for proper soluble expression of recombinant proteins (Ferrer et al. 2004). The N-terminal His-tagged *phr2*-pET28a construct was transformed into ArcticExpress™ with selection based on kanamycin, streptomycin and gentamycin resistance. The expression trial was set up according to the manufacturer's manual with both 0.5mM and 1mM [IPTG], and 4µl samples of soluble and insoluble fractions were loaded onto a 10% SDS-PAGE (Figures 5.22 and 5.23)

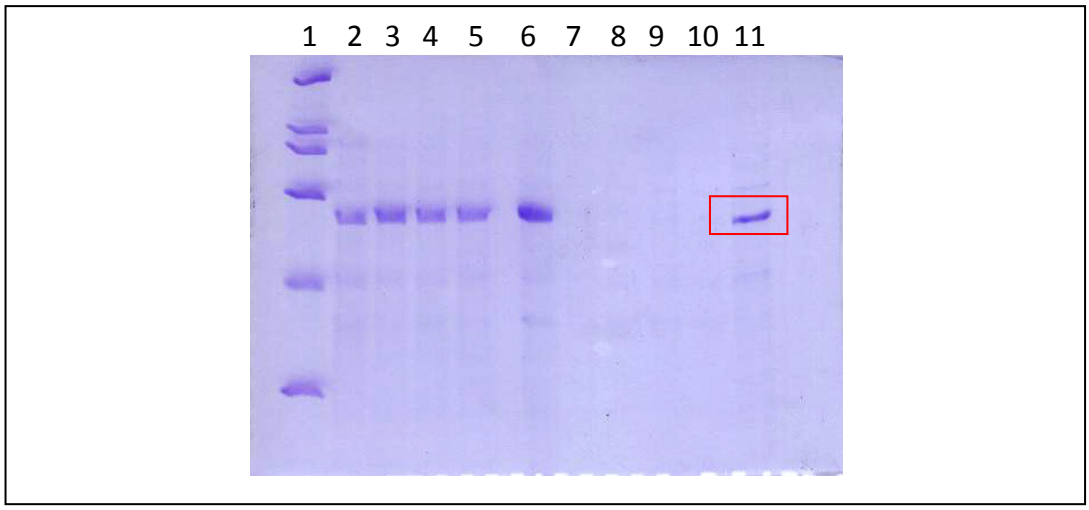


Figure 5.22: Protein expression analysis of his-tagged Phr2 protein using ArcticExpress™. Transformed ArcticExpress™ cells were induced with 0.5mM IPTG. 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11
Fraction	m	i	i	i	i	i	s	s	s	s	s
Time (hrs)	-	0	2	4	6	24	0	2	4	6	24

m = marker; **i** = insoluble; **s** = soluble

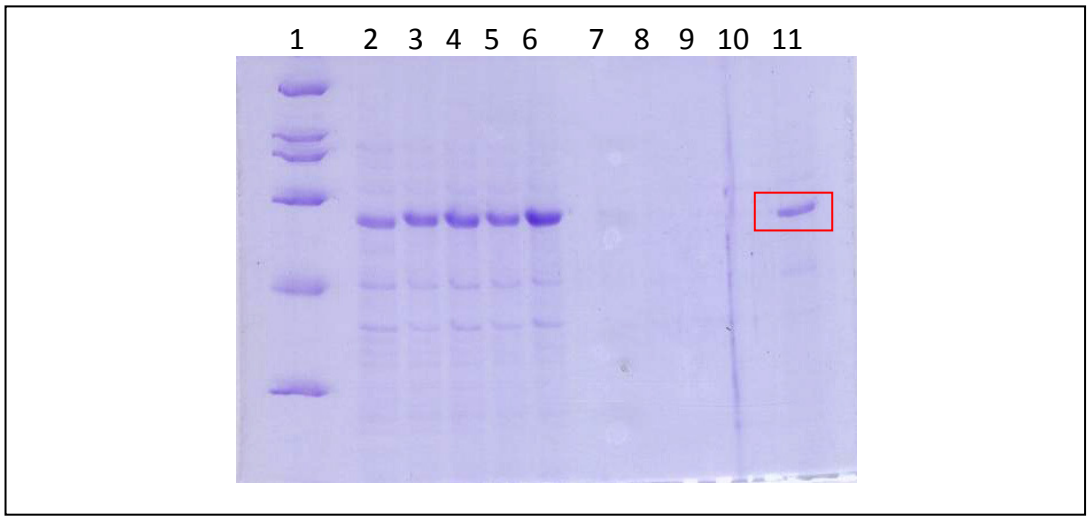


Figure 5.23: Protein expression analysis of his-tagged Phr2 protein using ArcticExpress™. Transformed ArcticExpress™ cells were induced with 1mM IPTG. 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11
Fraction	m	i	i	i	i	i	s	s	s	s	s
Time (hrs)	-	0	2	4	6	24	0	2	4	6	24

m = marker; **i** = insoluble; **s** = soluble

Figures 5.22 and 5.23 both show that the overall protein expression levels using ArcticExpress™ were much lower compared to expression in BL21 (DE3). However, the Phr2 protein, although mostly insoluble, appeared to be partially soluble at 24 hr (lane 11 in both figures). This was surprising, as previous expression trials have not shown a soluble Phr2 protein at 24 hr. A 250ml culture of the transformed ArcticExpress™ cells was induced with 0.5mM IPTG and incubated at 12°C for 24 hr. The soluble and insoluble fractions were prepared and the cell extracts were purified using the His•Bind column as described in Section 5.3.4. 5µl of the eluted fractions were run on a 10% SDS PAGE (Figure 5.24).

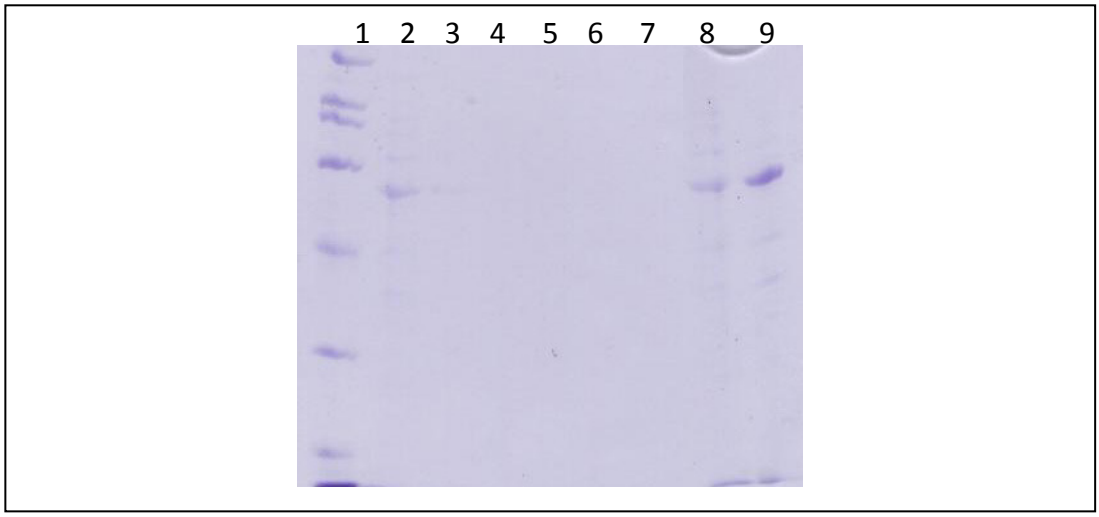


Figure 5.24: Analysis of eluted fractions following purification of cells grown in ArcticExpress™. 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9
Fraction	m	ft	wash	1st	2nd	3rd	4th	s	i

m = marker; **ft** = flow through; lanes **4** -**7** = eluted fractions; **s** =soluble; **i** = insoluble

Figure 5.24 shows that the protein band at approximately 60kDa was not eluted with an increasing concentration of imidazole, indicating that the soluble protein was, in fact, the Cpn60 protein, constitutively expressed from the pACYC-based plasmid. This explained why a band was observed in the 24 hr soluble fraction, as this Cpn60 is typically detected in SDS-PAGE analysis of ArcticExpress cell lysates at 56,900 Da.

Expression trials were conducted for the non-tagged *phr2*-pet-28a construct to assess whether the polyhistidine tag interfered with protein folding and structure. The C-terminal tagged construct was also used to assess whether the location of the tag played a potential role in protein insolubility. In parallel, expression trials for the three *phr1*-pet-28a constructs were performed to assess the solubility of this halophilic protein in a mesophilic expression host. All cells were induced with 0.5mM IPTG and incubated at 25°C.

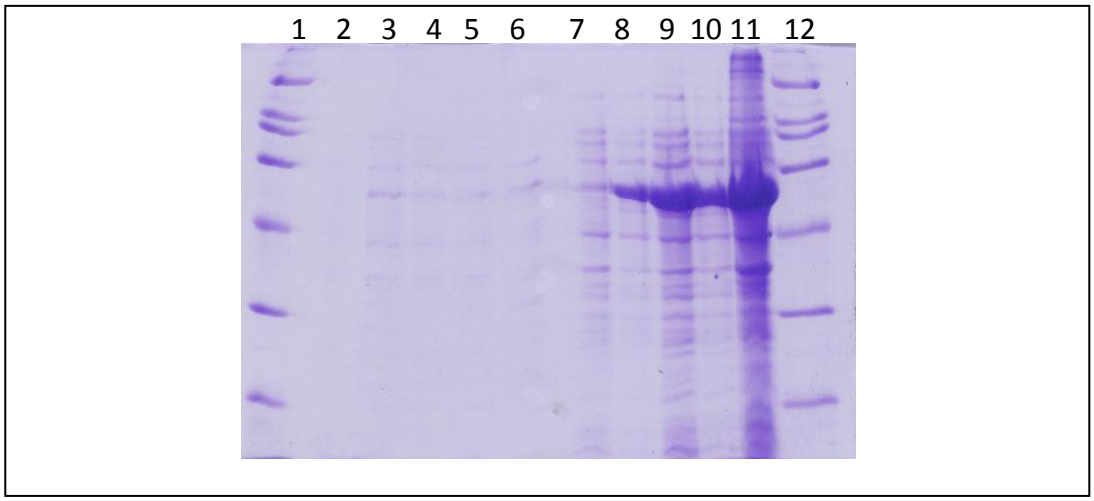


Figure 5.25: **Protein expression analysis of non-tagged Phr2 Protein.** 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11	12
Fraction	m	s	s	s	s	s	i	i	i	i	i	m
Time (hrs)	-	0	2	4	6	24	0	2	4	6	24	-

m = marker; **s** = soluble; **i** = insoluble

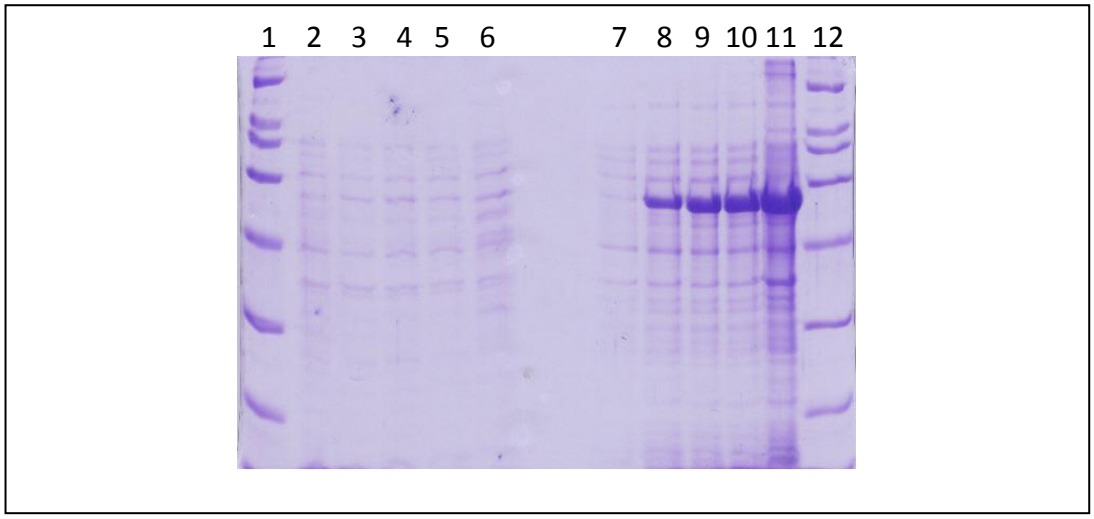


Figure 5.26: Protein expression analysis of C-terminal his tagged Phr2 protein. 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11	12
Fraction	m	s	s	s	s	s	i	i	i	i	i	m
Time (hrs)	-	0	2	4	6	24	0	2	4	6	24	-

m = marker; **s** = soluble; **i** = insoluble

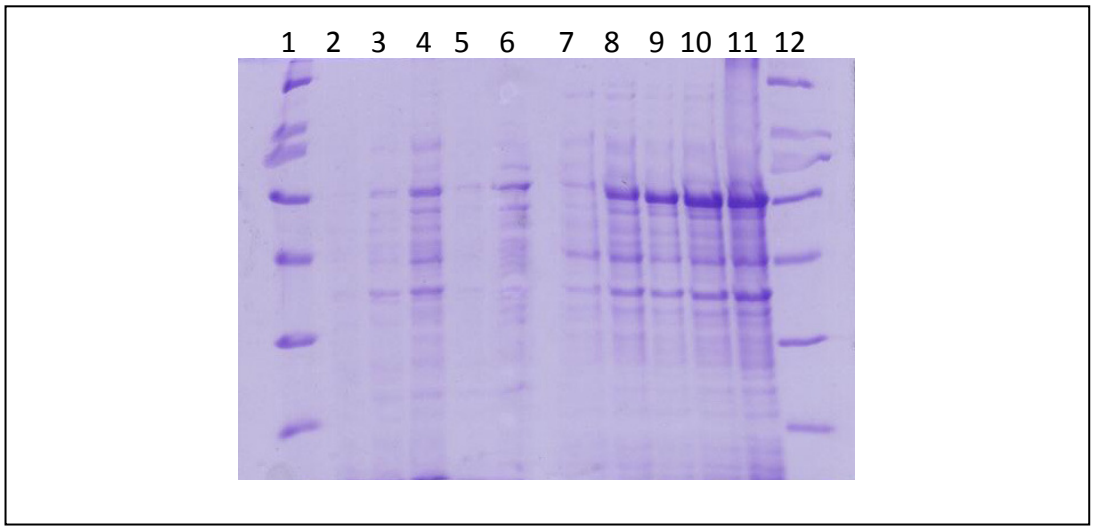


Figure 5.27: Protein expression analysis of non-tagged Phr1 protein. 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11	12
Fraction	m	s	s	s	s	s	i	i	i	i	i	m
Time (hrs)	-	0	2	4	6	24	0	2	4	6	24	-

m = marker; **s** = soluble; **i** = insoluble

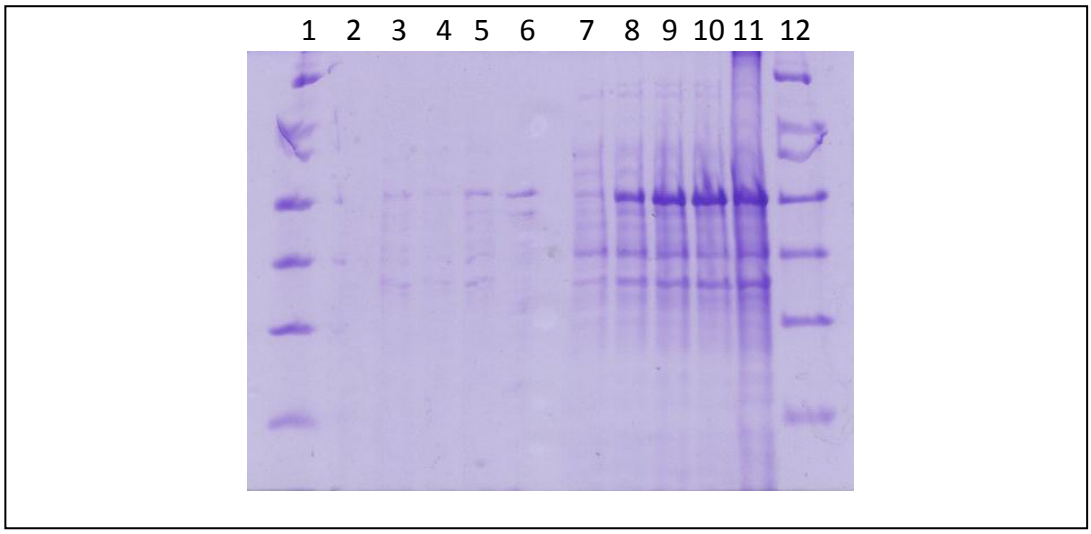


Figure 5.28: Protein expression analysis of C-terminal his tagged Phr1 protein. 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11	12
Fraction	m	s	s	s	s	s	i	i	i	i	i	m
Time (hrs)	-	0	2	4	6	24	0	2	4	6	24	-

m = marker; **s** = soluble; **i** = insoluble

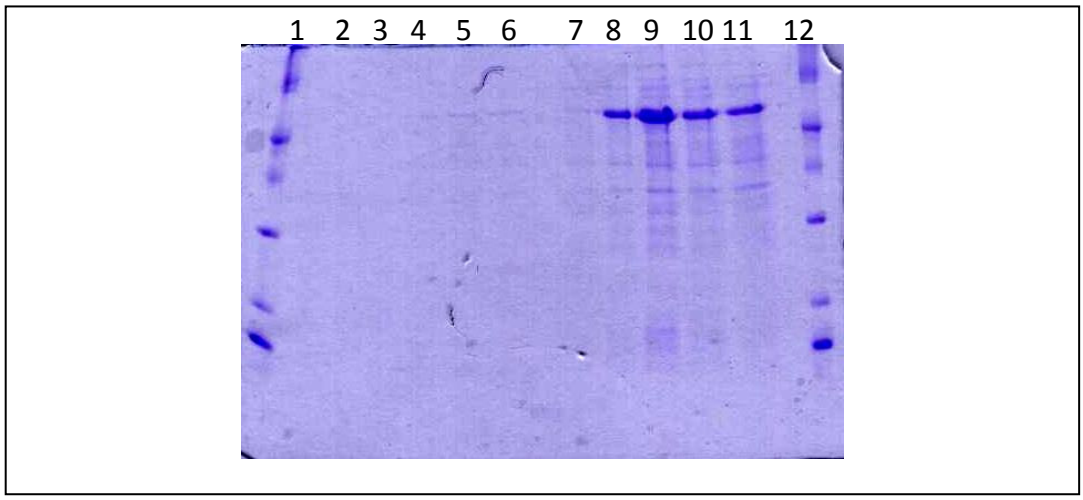


Figure 5.29: Protein expression analysis of N-terminal his tagged Phr1 protein. 4µl of each sample (soluble and insoluble fractions) was run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Lane	1	2	3	4	5	6	7	8	9	10	11	12
Fraction	m	s	s	s	s	s	i	i	i	i	i	m
Time (hrs)	-	0	2	4	6	24	0	2	4	6	24	-

m = marker; **s** = soluble; **i** = insoluble

It was evident from Figures 5.25 and 5.26 that protein solubility of Phr2 did not improve either when the His-tag was absent, or with a change in tag orientation. It was also evident from the insolubility of the Phr1 protein in Figures 5.27 to 5.29 that expression of halophilic proteins typically resulted in the formation of inclusion bodies when using a mesophilic expression host such as *E. coli* irrespective of the orientation of the His-tag.

5.4.4 Purification and Refolding

The large quantity of inclusion bodies formed was advantageous and desirable for the re-naturation or re-folding of the Phr2 protein. The inclusion bodies were first re-solubilized in 8M urea, followed by purification on the Ni charged column prior to re-naturation to increase the purity of the elutant. Due to the high concentration of the expressed protein in the insoluble fractions, the Phr2 protein often eluted in the unbound fractions before the addition of imidazole (Figure 5.30). This meant that the unbound fractions can be purified again to extract more Phr2 protein.

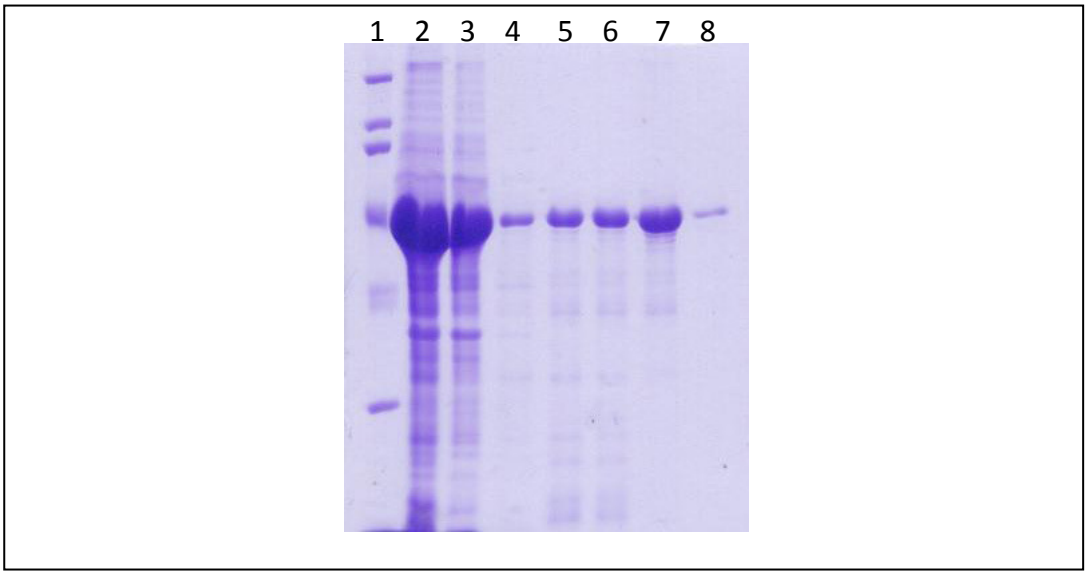


Figure 5.30: Purification of N-terminal His-tagged Phr2 (denaturing condition). 3ml of the insoluble fraction was purified on a Nickel column. 4µl of each sample was run on a 10% SDS-PAGE gel and stained with Coomassie Blue. Lanes 4 to 7 indicate increasing imidazole concentration up to 100mM.

Lane	1	2	3	4	5	6	7	8
Fraction	m	ft	wash	1st	2nd	3rd	4th	cw

m = marker; **ft** = flow through; **cw** = column wash

Figure 5.30 shows that the protein of the expected size was eluted in every fraction, most likely due to the high concentration of the expressed protein. This also showed that the amount of protein eluted was proportional to the concentration of imidazole used; therefore, subsequent batches of proteins were eluted with 100mM imidazole only.

Several re-folding protocols were consulted and tested to determine the best conditions to avoid protein loss due to precipitation. First trial involved dialysing an undiluted preparation of purified protein at room temperature on a stirrer using a buffer containing 4M NaCl, 1µM FAD and 5mM Triton-X; this resulted in precipitation after 5 hr on the stirrer. The concentration of protein was then decreased to 0.2 to 0.5mg/ml for the rest of the re-folding trials. Direct dilution was sought as an alternative to dialysis, where 2ml of the purified fraction was diluted 20-fold in re-naturation buffer, followed by dialysis to remove all components in the buffer except for 2M KCl. The presence of imidazole in the buffer, however,

interfered with the measurement of protein concentration, making the assessment of protein loss difficult. Step-wise dialysis was determined to be the most efficient method for maintaining a constant concentration of protein throughout the re-folding process. It was experimentally determined that the protein stayed in solution down to 0.2M urea, as precipitation occurred when urea was completely removed from the buffer. The presence of FAD was also problematic as it was not observed in any of the re-folded proteins. FAD has a characteristic absorption spectrum at 3 wavelengths: 378nm, 450nm and 467nm. Previous work on *E. coli* photolyase has shown that the FAD and the apoenzyme existed at a 1:1 stoichiometric ratio when purified (Sancar 2003). However, the re-natured protein did not appear to contain any FAD, as no peaks were detected in its respective spectra.

To check for the integrity of the re-folded protein, a 1ml protein sample was used to perform circular dichroism (CD) spectroscopy, courtesy of the Department of Chemistry, Queen Mary, University of London. CD can be used to determine whether a protein is folded and contains secondary structures in the far-UV spectral region (190-250nm). The chromophore at these wavelengths is the peptide bond which gives rise to a signal when it is in a folded environment (Whitmore and Wallace 2008). The CD scan as shown in Figure 5.31 demonstrates that the re-folded protein has secondary structure, with a higher percentage of α helixes to β sheets. The large peak shown was likely due to the high background of the salt-containing buffer used.

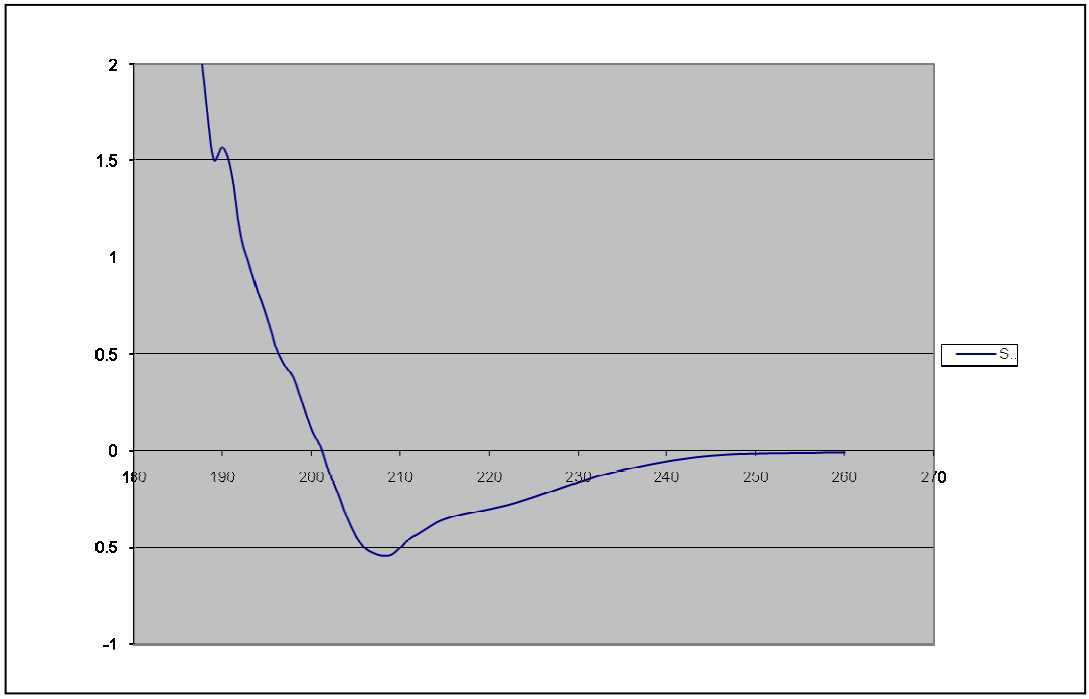


Figure 5.31: CD spectra of the re-folded Phr2 protein. Sample "s" indicated the protein used for the run. This signal is an average of the entire molecular population comprising of α-helix, β-sheet, and random coil structures.

5.5 Discussion

This section described the results of expressing the *Hfx. volcanii* Phr2 protein in a heterologous expression host, *E. coli*, in order to obtain high quantities of soluble, recombinant protein for the detection of enzymatic activity as well as attaining structural information on a halophilic photolyase. Successful heterologous expression of halophilic proteins has been shown to be a difficult task; in the cases of malate dehydrogenase from *Haloarcula marismortui* (Cendrin et al. 1993), dihydrofolate reductase from *Hfx. volcanii* (Blecher et al. 1993), HMG-CoA reductase from *Hfx. volcanii* (Bischoff and Rodwell 1997), citrate synthase and DHlipDH from *Hfx. volcanii* (Connaris et al. 1999) and signal peptidase from *Hfx. volcanii* (Fine et al. 2006), all proteins produced were either soluble yet inactive, or insoluble inclusion bodies that required further activation by salt or re-naturation, respectively. Successful cases of active and soluble expression of halophlic proteins from *E. coli*, such as the glycolic pathway enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH) from *Hbt. salinarum* (Ha et al. 2007), are few and far between. Nevertheless, the yields of heterologously expressed proteins from *E. coli*

were, in many cases, higher than those expressed from the native organism (Connaris et al. 1999), as demonstrated by the low yield of homologously expressed Phr2 shown in Chapter 2. The high yield of recombinant proteins produced would allow for easier purification and downstream applications such as functional assays and structural analysis.

Both *phr1* and *phr2* were PCR amplified and cloned into appropriate intermediate vectors before the insertion into an expression vector. The gene sequences were checked for errors and the inserts with the correct sequences were ligated into pET-28a. The clones were transformed into BL21 (DE3) to look for the expression of the Phr proteins and to conduct initial small-scale expression trials. SDS-PAGE revealed that for small-scale expression the Phr2 protein appeared soluble; first time attempt to purify the protein showed that a much higher concentration of soluble protein was required for purification. It was found, however, that the protein became insoluble when a larger culture volume was induced. This was tested on numerous occasions using different combinations of temperature, induction time and IPTG concentrations. All trials showed consistently that the protein remained insoluble.

A C-terminal His-tagged *phr2*-pET28-a construct as well as a non-tagged *phr2*-pET28a construct were also made to assess whether this affinity tag interfered with protein solubility, as previous research has shown that an N-terminal His-tag interfered with the solubility of an intracellular binding protein expressed in *E. coli* where the resulting protein was largely insoluble and unstable. However, removal of the tag during re-folding resulted in a soluble protein exhibiting desired folding characteristics and which was therefore purified (Ramage et al. 2002). SDS-PAGE showed that the Phr2 protein was, again, largely insoluble irrespective of the location or presence of the tag. Attempts at purifying the insignificant amount of soluble protein proved unsuccessful. The *phr2*-pET28a construct was transformed into Rosetta™ and ArcticExpress™ strains to see whether a change in host strain could increase the solubility of the Phr2 protein. Neither strain resulted in an increase in protein solubility. The large quantity of insoluble protein was subjected to re-folding treatments to re-gain formation of the protein native structure.

The formation of inclusion bodies, or insoluble aggregates, is a result of the accumulation of folding intermediates rather than the native or unfolded polypeptides. It occurs when the rate of translation exceeds the rate of folding of the nascent polypeptide to the native conformation (Kane and Hartley 1988; De Bernardez-Clark and Georgiou 1991). The routine use of strong promoters and high inducer concentrations often lead to an excess of product yield compared to the total cellular protein. Under these conditions, the rate of protein aggregation exceeds proper folding where folding modulators cannot interact with the protein quick enough. Nevertheless, this inactive protein in the form of inclusion bodies offers several advantages to proteins of soluble form, such as a high degree of purity in the target protein, as well as the resistance to proteolytic degradation compared to the soluble fractions. Furthermore, the halophilic proteins can be selectively purified from the contaminating host protein due to the high salinity of the purifying buffer, thus retaining the near native environment of these halophilic proteins. Lastly, these inclusion bodies can be re-solubilized and the activity of the protein can be re-gained through established methods of re-folding (Tsumoto et al. 2003).

The inclusion bodies obtained from the expression of *Hfx. volcanii* Phr2 were first solubilized, followed by purification on a Ni affinity column, as numerous reports have shown that higher refolding yields were achieved when the re-solubilized inclusion body proteins are purified prior to refolding treatments (Tran-Moseman et al. 1999). The purified fractions were then transferred into low denaturant conditions to allow for formation of the native structure. The first technique used was dialysis, where the change from the denaturing to the native buffer conditions was gradual. The protein was dialysed against buffers of decreased denaturant concentrations to allow formation of native structure. This, however, caused aggregation just after 5 hr. In dialysis, it has been found that aggregation occurred often due to the increased number of folding intermediates (Gu et al. 2002). To avoid this problem, direct dilution was utilized, where the protein in the strong denaturant condition was diluted into a refolding buffer in a drop-wise fashion. Although aggregation did not occur, it was difficult to measure protein concentration due to the presence of imidazole which affected true protein read on the spectrophotometer. Imidazole absorbs UV radiation at 280nm, meaning that a

sample measured at 280nm shows an increase in absorbance above the background, over estimating the concentration of protein. The same problem occurred when Bradford assays were used, as the presence of urea interfered with protein reading, which again gave a high background reading (Bradford 1976; van Kley and Hale 1977). The BCA protein determination method was not suitable as the presence of disulfide reducing agents, including DTT, interferes with the assay by artificially increasing the colour intensity, therefore giving an inaccurate protein read (Brown et al. 1989).

Step-wise dialysis was finally determined as the most accurate method for quantification of protein post re-folding, as the final buffer contained only salt and a low amount of urea, neither of which affected the protein quantification methods used. In this dialysis system, descending concentration of denaturant was used where the unfolded protein sample was slowly brought to equilibrium with the buffer at each stage of dialysis (Tsumoto et al. 2003). The protein sample was first dialyzed against the renaturation buffer containing an excess amount of FAD as described in Section 4.3.5, overnight at 4°C. The next day the protein was transferred to fresh buffer containing 2M urea and 2M KCl, again incubated overnight at 4°C. Finally, the protein was transferred to a solution containing 0.2M urea and 2M KCl, as it was experimentally determined that below 0.2M urea aggregation occurred. This was not surprising, as it has been shown that low residual concentrations of denaturant improved refolding yields by suppressing aggregation (De Bernardez Clark et al. 1999). Overall, step wise dialysis proved to be a good system for re-folding the Phr2 protein where minimal loss of protein occurred, and that protein remained in solution after re-folding. While the CD spectrum suggested that the Phr2 protein had secondary structures, it cannot predict whether it was folded correctly. As FAD was not incorporated into the protein, it was a likely indication that the residues important for binding the chromophore may not have been in the right orientation.

Although it was not possible to obtain a soluble Phr2 protein at this stage, the mass production of inclusion bodies offered high purity and quantity as opposed to expression from a homologous strain where no distinct over-expression of Phr2

protein was detectable. This also proved that the soluble expression of halophilic proteins in *E. coli* was difficult to achieve, where the low ionic strength internal environment possibly prevented the correct folding of halophilic enzymes. Given the time constraint, it was not feasible to look into other approaches to overcome this insolubility problem. One such approach would be to fuse an aggregation-prone polypeptide to a highly soluble partner such as the maltose-binding protein (MBP) (Pryor and Leiting 1997), glutathione s-transferase (GST) (Nygren et al. 1994), and thioredoxin (TRX) (LaVallie et al. 1993), all of which have been shown to promote the proper folding of the attached protein into its biologically active conformation. The GST fusion tag has been used successfully with an acetyltransferase protein, Vgn0487h, from *Hbt. Salinarum*, where enough fusion proteins were observed in the soluble fraction for a pull-down purification of the protein (Choi et al. 2005). Regardless of the insolubility problem, the re-natured protein was shown to contain secondary structure, an indication that the protein was folded to give such conformation. The information that is lacking pertains to whether this protein is properly folded in its native form, and whether it exhibits desired activity. In the next chapter, an immunoassay set up to test the activity of the re-folded protein is described.

Functional Characterization of Phr2

6.1 Introduction

Repair of UV-induced pyrimidine dimers is dependent on a highly efficient light-dependent photoreactivation mechanism, where haloarchaea are known to be highly tolerant to UV in their natural habitats containing visible light (Hescox and Carlberg 1972; Fitt et al. 1983; McCready 1996; McCready and Marcello 2003). In *Hfx. volcanii*, the mechanisms underlying its high tolerance to UV remain unknown. Genome analysis of a model haloarchaeon, *Halobacterium* sp. NRC-1, has identified two putative photolyase homologues, *phr1* and *phr2*, and deletion mutant analysis revealed that only *phr2* was required for UV-induced DNA repair during recovery in visible light, suggesting that *phr2* encoded a functional photolyase (McCready and Marcello 2003; Baliga et al. 2004). In addition to the light dependent photoreactivation, *Halobacterium* sp. NRC-1 also possesses homologs of the bacterial *uvrA*, *uvrB* and *uvrC* genes, shown to be required for the removal of UV damage in the absence of photoreactivating light. For many years, it was thought that *Halobacterium* does not possess a dark repair system as demonstrated by Grey and Fitt, where no viable cells were found in UV irradiated cultures of *Hbt. cutirubrum* photoreactivating light with no exposure to photoreactivating light (Grey and Fitt 1976). However, in a more recent study by McCready et al., deletion mutants lacking functional *uvrA*, *uvrB* or *uvrC* genes, as well as a *uvrA uvrC* double mutant, were sensitive to UV and unable to remove CPD or 6-4 photoproducts from their DNA post UV irradiation in the absence of photoreactivating light. Upon incubation with visible light, however, the sensitivity was attenuated, placing an emphasis on the highly efficient photoreactivation in *Halobacterium*. Interestingly, these deletion mutants are substantially less UV sensitive than excision repair mutants of *E. coli* or yeast, suggesting for the existence of multiple tolerance mechanisms within the genome (Crowley et al. 2006). *Hfx. volcanii* has also been shown to excise UV lesions; however, this remains a largely uncharacterized repair process (McCready 1996).

Results from the current study have identified novel photolyase-like genes, *phr1* and *phr2*, for the first time in *Hfx. volcanii*. Predicted structural analysis of the Phr2

protein suggests that, unlike bacterial photolyases, halophilic photolyases use 8-HDF as its light-harvesting cofactor (Chapter 2). Homologous expression of the Phr2 protein in the native *Haloferax* strain were too low (Chapter 3) and heterologous expression in *E. coli* showed that the expressed protein was mostly insoluble and required re-folding treatments (Chapters 4 and 5). However, for both the homologously and heterologously expressed Phr2 proteins, activity was not assessed. This chapter demonstrates the functional activity of the expressed Phr1 and Phr2 proteins using both *in vivo* and *in vitro* assays. An *in vivo* UV survival assay was developed to assess the photosensitivity of the homologous expression strains to UV and the repair by the expressed proteins upon exposure to white light illumination. This assay allowed for clear visual qualitative evidence of photorepair. *In vitro* assays included a) the cell viability assay used to quantify survival post UV irradiation by the measurement of the release of ATP and b) an antibody-based assay optimized for use with the Phr2, Phr1 and E3 overexpression strains as well as the re-folded Phr2 protein prepared in Chapter 5.

6.2 Materials

6.2.1 Cell culture

Yeast extract, peptone and NaCl were supplied by Fisher Scientific, Loughborough, UK. Agar, ampicillin, carbenicillin, tryptone, polyethylene glycol (PEG, Average MW 600) and thymidine were from Sigma-Aldrich, Gillingham, UK. Bacto Yeast Extract, Bacto Casamino Acids and Bacto Agar were supplied by BD-Biosciences, Oxford, UK. Bacteriological peptone was obtained from Oxoid, Basingstoke, UK. Unless specified otherwise, all other reagents were ordered from Sigma-Aldrich, Gillingham, UK.

6.2.2 UV survival assay

UVILite Ultraviolet Filtered Lamps (wavelength: 254nm, intensity at 15cm $\mu W/cm^2$: 710) were purchase from VWR Jencons. Fluorescent lamps (white light source) were obtained from Philips (15W; emission: 400-700nm). UV output was measured using a Latarjet radiometer.

6.2.3 Cell viability assay

Components for the BacTiter-Glo Microbial Cell Viability Assay were obtained from Promega.

6.2.4 Immunoassay

Primary anti-thymine dimer mouse monoclonal antibody (clone KTM53) was purchased from Kamiya Biomedical Company, USA. Peroxidase conjugated anti-mouse antibody was obtained from Sigma-Aldrich, Gillingham, UK. *E. coli* photolyase was obtained from Trevigen, USA. Nitrocellulose membrane (0.2μm) and herring sperm DNA (degraded) were also obtained from Sigma-Aldrich, Gillingham, UK. The x-ray film was obtained from Fuji, Japan.

6.3 Methods

6.3.1 UV survival assay

Cells were grown in HV-YPC for 3 days until an OD_{600} greater than 1.2, and 50μl of the cultures were streaked unto HV-YPC agar plates. The plates were incubated for 3 hr at 42°C, followed by exposure to a germicidal UV lamp (emitting mainly at 254nm) placed at a distance of 30 cm above the plates for up to 5 min at a UV output of 1.15 $J/m^2/s$ in the dark. A piece of glass sheet wrapped with foil was used to cover the plates to allow different times of exposure to UV. Following UV exposure half of the plates were exposed to a white light source (fluorescent lamp 15W; emission: 400-700nm) for 30 min, after which all plates were wrapped in foil and incubated at 42°C for up to 3 days.

6.3.2 Cell viability assay

The BacTiter-Glo Microbial Cell Viability Assay was used to determine the number of viable cells by a culture-based quantitation of the amount of ATP released. Both the BacTiter-Glo Buffer and the lyophilized BacTiter-Glo Substrate were equilibrated to room temperature before use, followed by mixing the two reagents together to reconstitute the lyophilized enzyme/substrate mixture. The reagents were then again left at room temperature for 15 min before use. Two homologous expression strains were used for this assay: Phr2 and E3, where 10ml cultures of

each strain were grown for 3 days, followed by a 100-fold dilution in the same growth media. Liquid cultures were prepared as described in Section 2.3.6. 100µl of the diluted cultures were dispensed into two white Greiner 96-F bottom plates. The plates were UV irradiated for up to 4 min (UV output 1.15 J/m^2/s), after which the "light" designated plate was exposed to a white light source for 30 min and the "dark" designated plate was wrapped in foil and left at room temperature for the same amount of time. Both plates were then incubated in the dark at 45°C for 18 hr. In the darkroom, 40µl of the incubate cultures was aliquoted into a new opaque 96 well plate, followed by the addition of an equal volume of the re-constituted substrate mixture into each well to allow for ATP bioluminiescence detection.

6.3.2.1 Detection of luminescence

The plates were then read on the Fluostar Omega Multifunctional Microplate Reader (BMG labtech) to record luminescence using the emission filter 480-10 (No.1A) with a gain setting at 3600. The measurement interval time was set at 0.24s and the reading was performed at room temperature. The plates were covered in foil prior the actual read to avoid exposure to light. Each experimental condition had 8 replicates for each strain and the values represent the means from 3 independent experiments. Values were normalized to 100% against un-irradiated controls independently for the dark and the light conditions. The variations in values within and between each trial were taken into consideration when collating all 3 data sets together and generating the results represented in a graphic form.

6.3.2.2 Statistical analysis

Data presented as means ± SEM of 4 independent cultures. Statistical significance was determined using two-way ANOVA (post hoc: Bonferroni post test).

6.3.3 Immunoassay

Herring sperm DNA was dissolved at 5mg/ml in sterile deionized water and left to dissolve overnight at 4°C. The DNA solution was further diluted to 0.050mg/ml or 0.025mg/ml and 50µl was dispensed into wells of a white, opaque 96-well plate. The plate containing DNA was UV-irradiated up to 3 min (UV output 1.15 J/m^2/s) in

the dark. 3µl of the irradiated DNA was dispensed unto a 0.20µm nitrocellulose membrane. The membrane was left to dry at room temperature for one hr, followed by neutralizing in 2X SSC (0.03M Sodium Citrate, 0.3M NaCl, 0.1mM EDTA pH 7.00. The membrane was baked in the oven at 80°C for 150 min to fix the DNA onto the membrane; it was then stored at 4°C in TBS-T (10mM Tris-HCl pH 8.0, 0.1% Tween-20, 200mM NaCl) for further use.

The dried membranes were incubated with either the purified and re-natured Phr2 protein or prepared lysates from H98 Phr2 and Phr1 expression strains. The lysates were prepared by harvesting the cells at 8000rpm for 15 min, and re-suspending the pellet in 5ml extracting buffer (2M NaCl, 50mM Tris HCl pH8.0, and 1mM EDTA). The re-suspension was sonicated at 5 20s bursts (amplitude = 14 micron) in an ice bath with 40s cooling in between each burst. 200µg of the re-natured Phr2 protein was made up to 2ml in the same storage buffer it was in. 2ml of prepared lysates of both Phr2 and Phr1 transformed strains were used where the concentration factor not taken into consideration, as it was not feasible to measure protein concentration as stated in Chapter 3. Membranes were incubated either in the dark or under a white light source for 2 hr with the membranes covered to avoid desiccation. Membranes were also incubated in water as a control. Post incubation, the membranes were rinsed with TBS-T buffer twice, followed by blocking in 10% milk/TBS-T for 1 hr at room temperature. The membranes were then incubated with the mouse monoclonal antibody raised against thymine dimers, diluted 1:500 in 5% milk/TBS-T, for 1 hr. Following a 4 x 10 min wash in TBS-T buffer , the membranes were incubated with a peroxidase-conjugated anti-mouse antibody, diluted 1:4000, for 1 hr. The membranes were again washed 4 x 10 min in TBS-T before overlaying the chemiluminescent HRP substrate (detection agent) on top, DNA side up, for 5 min. The UV-damaged DNA was detected using anti-CPD monoclonal antibody, which was bound to the secondary peroxidase-conjugated antibody. The peroxide catalyzed the oxidation of luminal, where once oxidized, the luminal emitted light as it decayed to its ground state. The amount of DNA damage recognized by the anti-CPD monoclonal antibody was proportional to the intensity of each dot. With increasing levels of UV-irradiation (at a constant DNA concentration), more damage would be induced; hence, the dots became progressively darker. The

chemiluminescence detection protocol was performed according to the manufacturer's instructions. To evaluate antibody binding, the blots were first transferred to an x-ray film and developed using Optimax X-Ray Film Processor (Xograph Imaging System) in the dark room.

6.3.3.1 Quantitation of dimers

To quantitate the amount of dimers present, a dot blot analysis tool from LabWorks was used to calculate the integrated optical density (IOD) of each dot. The IOD is a measure of the summation of the grey levels of all pixels in any single dot. The intensity of the dot was proportional to the amount of damage present in that particular dot. Time zero dots were designated the background dots, and all other IOD values obtained from the rest of the dots were normalized against the pooled background value.

6.3.3.2 Statistical analysis

Data presented as means ± SEM of 4 groups (culture or protein) from 3 independent experiments (unless otherwise stated). Statistical significance was determined using two-way ANOVA (post hoc: Bonferroni post test).

6.4 Results

6.4.1 UV survival assay

Light and dark DNA repair activities were assessed for both Phr2 and Phr1 proteins expressed in the homologous H98 strain. The negative control was an over-expressing strain for the E3 gene, which encodes dihydrolipoamide dehydrogenase, an enzyme involved in central metabolism and with no known function in photoreactivation. The His-Phr2 strain was used to assess the effect of the His-tag on the function of the Phr2 protein. 50µl of each culture was sufficient to produce a longitudinal strip when streaked on a plate, as seen in Figure 6.1. This assay was optimized to produce clear and visible results to demonstrate each strain's ability to carry out DNA repair and grow after exposure to UV-C at a UV dose of 1.15 $J/m^2/s$. The plates were then given 30 min of white light exposure to activate the photolyase, after which all plates were incubated at 45°C in the dark for 3 days. The length of

the strip corresponded to increasing UV exposure from the top to the bottom of the plate; this was used as an indication of cell survival.

The Phr2 expression strain was resistant to UV up to a UV dose of 138 J/m^2, whereas for the Phr1 expression strain resistance was seen up to only 69 J/m^2 (Figure 6.1a). In Figure 6.1b, the E3 and the Phr1 expression strains were comparable to each other, maintaining the same level of UV resistance up to 69 J/m^2. In Figure 6.1c, the His-Phr2 expression strain was seen to be more UV resistant compared (UV dose 207 J/m^2) to the E3 expression strain (69 J/m^2). In Figure 6.1d, no difference was observed between the Phr2 and His-Phr2 expression strains; both strains were equally resistant to UV up to a UV dose of 207 J/m^2. This provided an indication that the presence of the His-tag did not interfere with the Phr2 protein function. Figures 6.1a and 6.1c clearly demonstrated that the Phr2 expression strain, in the presence of white light post UV irradiation, resulted in the strain's superior resiliency to UV over both the control E3 and the Phr1 expression strains. Looking at 6.1b, it was also apparent that Phr1 had the same UV sensitivity as the control strain.

Identical plates incubated in the absence of white light showed that all expression strains were equally resistant to UV up to a UV dose of 69 J/m^2 (Figures 6.2 a-d), an indication that the dark repair mechanism was maintained at the same rate of efficiency for every strain regardless of the protein expressed. This also demonstrated that white light was a crucial factor in the enhanced survival of the strain expressing the Phr2 protein, as this strain was more UV resistant when incubated in the presence of white light. The images shown in Figures 6.1 and 6.2 are representative of 3 independent sets of experiments done on separate batches of *Hfx. volcanii* cultures.

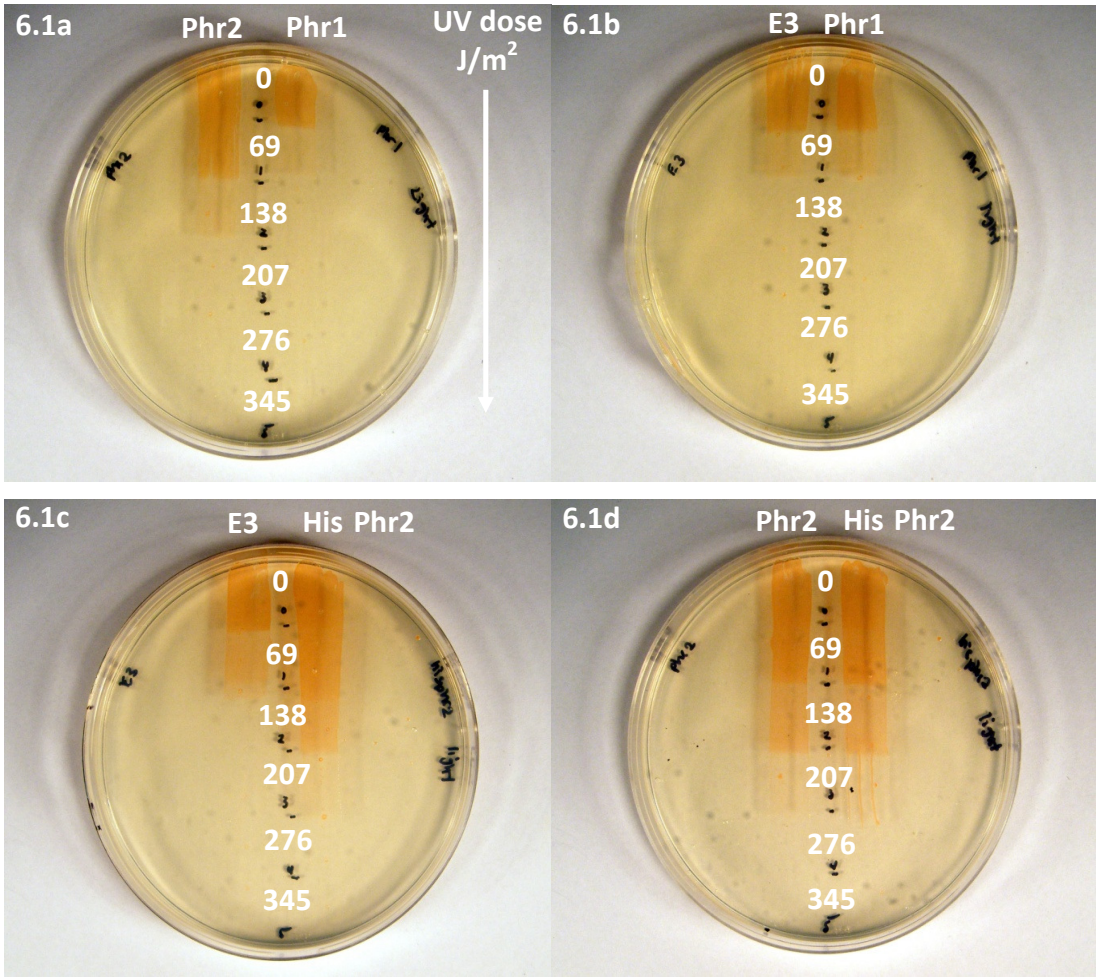


Figure 6.1a to 6.1e: Survival of *Haloferax* overexpression strains exposed to UV followed by white light. The UV doses were as indicated on the legend, with a UV output at 1.15 J/m^2/s and exposure up to 5 min. The strains were grown for 3 days, where 50µl of the cultures were streaked onto the plates vertically down the centre of each plate. The plates were UV-irradiated and exposed to white light for 30 min. These results were visualized after 3 days of incubation in the dark at 45 °C. **6.1a:** Phr2 vs. Phr1; **6.1b:** E3 vs. Phr1; **6.1c:** E3 vs. His-Phr2; **6.1d**: Phr2 vs His-Phr2

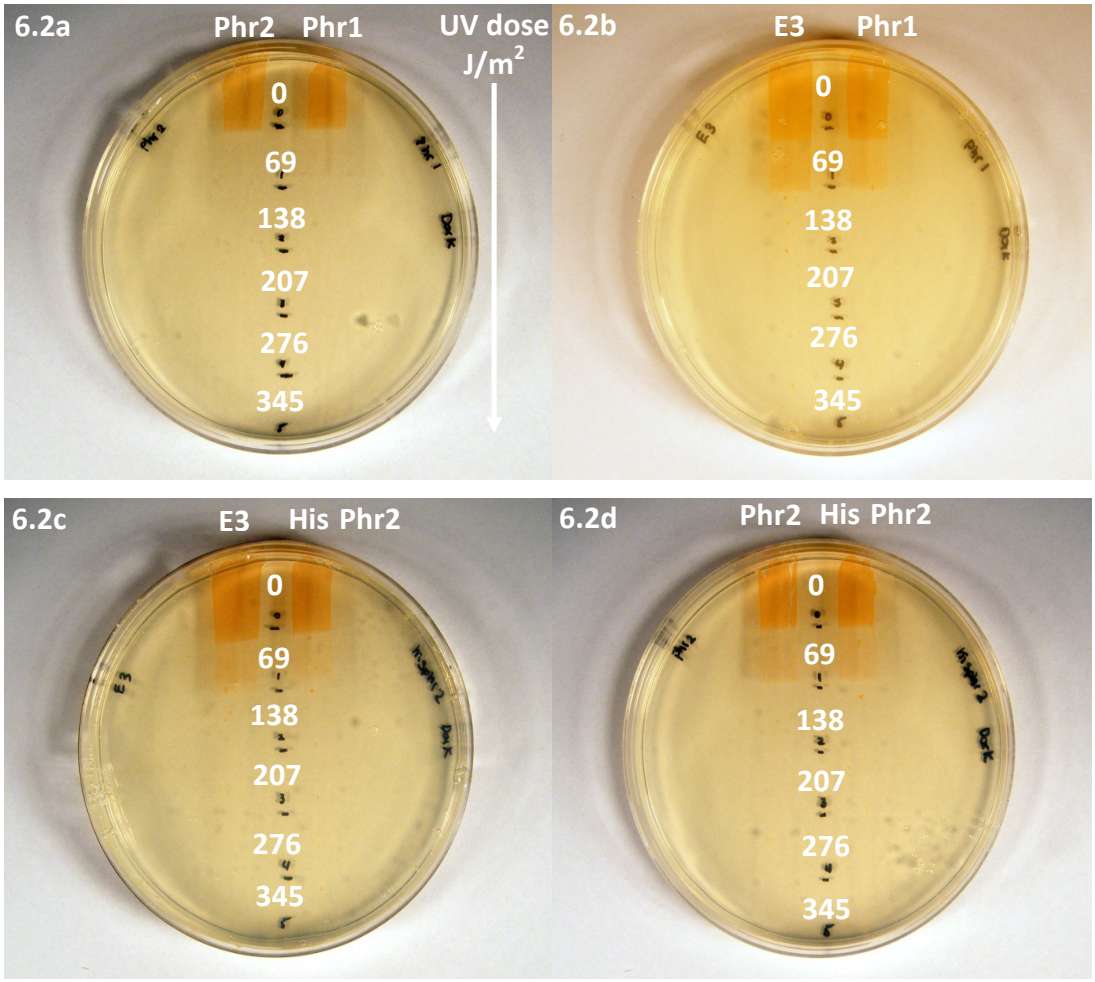


Figure 6.2a to 6.2e: **Survival of *Haloferax* overexpression strains exposed to UV and no white light.** The UV doses were as indicated on the legend, with a UV output at 1.15 J/m^2/s and exposure up to 5 min. The strains were grown for 3 days, where 50µl of the cultures were streaked onto the plates vertically down the centre of each plate. The plates were UV-irradiated and kept in the dark immediately post UV exposure. These results were visualized after 3 days of incubation in the dark at 45 °C. **6.2a:** Phr2 vs. Phr1, **6.2b:** E3 vs. Phr1, **6.2c:** E3 vs. His-Phr2, **6.2d:** Phr2 vs. His-Phr2

6.4.2 Cell viability assay

The BacTiter-Glo assay system was utilized for a quantitative analysis of the effects of white-light-induced survival of cells exposed to UV. Firefly luciferase, purified from *Photuris pennsylvanica,* was used in this system because of its stability in ATP assays. The results were presented as % ATP release through RLU (relative luminescence units) by the cells; thus, cell viability can be interpreted as a correlation of the measured ATP levels.

Three major observations can be made with regard to Figure 6.3 showing % ATP release of each strain at various UV exposure times. Firstly, the "light" Phr2 expression strain resulted in the highest overall % ATP levels throughout the experiment, correlating to a higher % survival for cells of this strain. A post test comparing the means of ± SEM for the "light" Phr2 and "light" E3 indicated that a significant difference was observed at a UV dose of 276 J/m^2 ($P < 0.001$). The post test has also indicated that the differences between "light" Phr2 and "dark" Phr2 were significant at all UV doses ($P < 0.001$).

The second trend observed is the gradual decrease in % ATP levels for the "light" E3, "dark" E3 and "dark" Phr2 strains with increasing UV exposure, correlating to decreased cell viabilities. A post test comparing the means of ± SEM for these three groups indicated that no significant differences occurred between the "light" E3 and the "dark" E3 strains at all UV doses. Also, no significant differences were found for the "dark" E3 and "dark" Phr2 strains at all UV doses ($P < 0.001$) with an exception at a UV dose of 207 J/m^2, where the "dark" E3 strain experienced a slight increase in % ATP.

The third observation gathered from Figure 6.3 is that a prolonged UV exposure is required to reach a median LD_{50} (dose required for a decrease in survival of half of the tested population). At a UV dose of 345 J/m^2, % ATP for light E3, dark Phr2 and dark E3 were 80.77, 73.92, 77.18, respectively, indicating a decrease in cell viability between 20 to 27%, approximately half of what is needed to reach the desired LD_{50}.

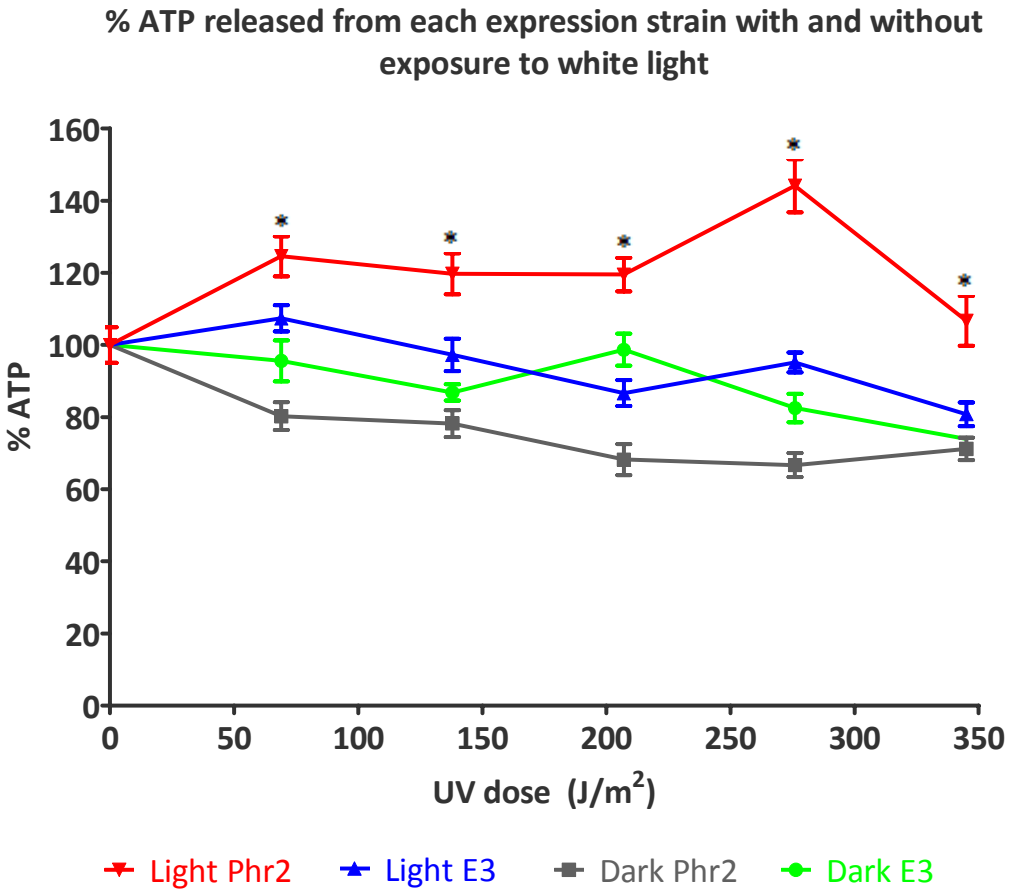


Figure 6.3: % ATP released for Phr2 and E3 expression strains. % ATP was plotted against increasing UV doses (UV output 1.15 J/m^2/s) under both light and dark conditions. Phr2 and E3 expression strains were grown for 3 days, followed by a 100-fold dilution in media. The diluted cultures were dispensed into two white plates, followed by UV irradiation. Half of the plates were exposed to a white light source for 30 min and other half were kept in the dark. Both plates were then incubated in the dark at 45° C for 18 hr. 40µl of the incubate cultures was aliquoted into a new opaque 96 well plate, followed by the addition of an equal volume of the re-constituted substrate mixture into each well for luminescence detection. Values are presented as means ± SEM from 4 independent cultures, where each group contained 24 replicates per condition. *Statistical significance at $P < 0.001$ compared with the "light" E3 and dark "Phr2."

6.4.3 Immunoassay

The dot blot immunoassay was used to measure the repair of total photoproducts induced in DNA by short-wave ultraviolet light. This method was adapted from the one previously described by McCready (McCready 1999) with modifications based on the method described by Storey (Giampieri and Storey 2004). In this assay, herring sperm DNA solution was subjected to treatments with UV irradiation and white light, followed by hybridization onto a membrane and allowed to air dry, followed by neutralization in 2X SSC and baking at 80°C to fix DNA unto the membrane. It was crucial, firstly, to determine the type of membrane to be used for the hybridization of DNA. Nylon membranes Hybond-N+ and Nybond-N, as well as nitrocellulose membranes from both Sigma and VWR were tested; the nitrocellulose membrane from Sigma, by far, allowed for the best hybridization of DNA and so was chosen for the rest of the experiments. The second adjustment to the assay was determining the amount of DNA to be fixed unto the nitrocellulose membrane. Serial dilutions of herring sperm DNA at 5mg/ml were performed in a 96 well plate followed by UV irradiating the plate for up to a UV dose of 345 J/m^2. 3µl of the irradiated DNA was blotted unto the membrane, where the rest of the experiment was carried out as outlined in Section 6.3.3. The scanned x-ray film was as shown in Figure 6.4.

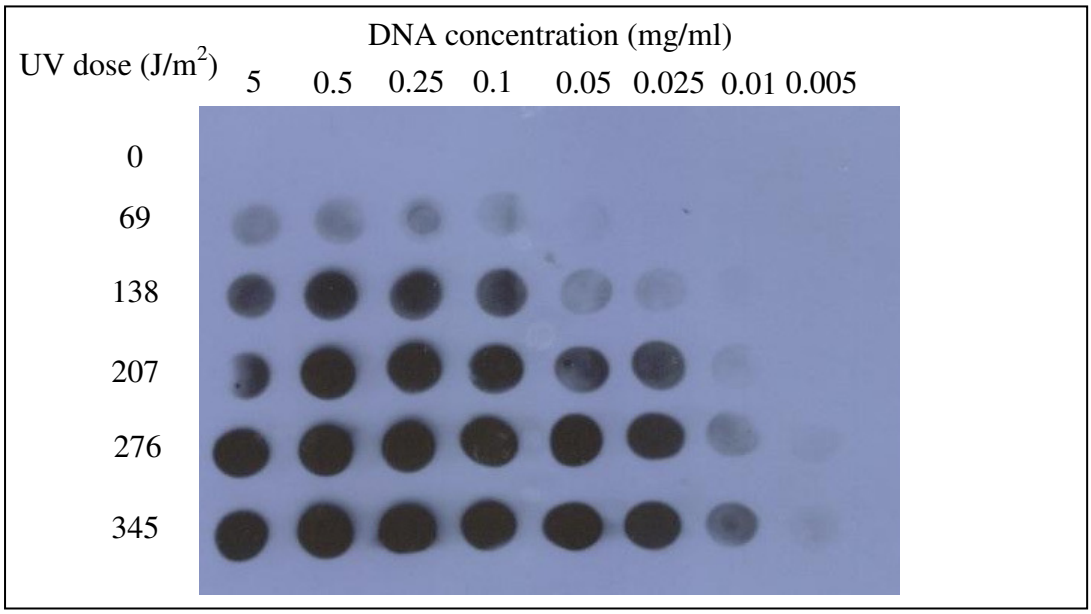


Figure 6.4: Effect of UV exposure on CPD lesions in DNA . The immuno-dot blot of varying DNA concentrations exposed to UV. Rows indicate DNA concentrations (5mg/ml stock). Column indicates the increase in UV dose in J/m^2. The figure is a representative image and its corresponding quantitative graph is shown in Fig 6.5.

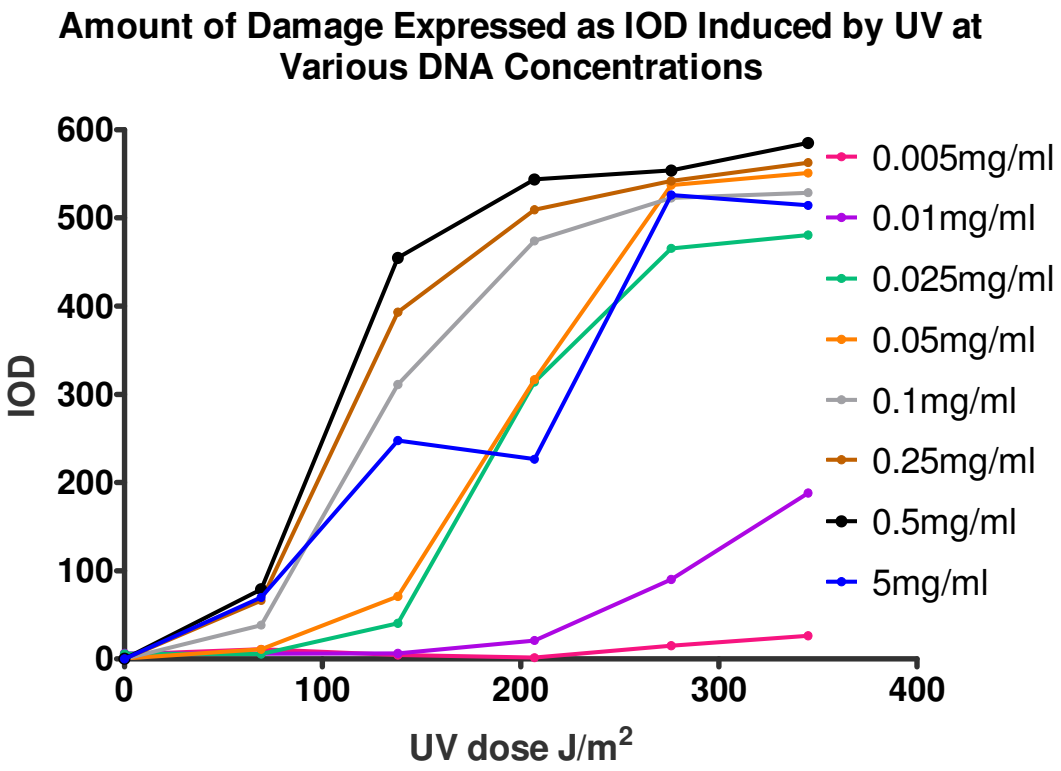


Figure 6.5: Effect of UV exposure on CPD lesions in increasing concentrations of DNA concentrations. The graph represents the quantitation of the immuno-dot blot data shown in Figure 6.4. IOD values (Relative units) were calculated against controls (water alone). Figure legend indicates increasing DNA concentrations.

As shown in Figure 6.4, DNA concentrations at 0.05mg/ml and 0.025mg/ml gave clear, incremental amounts of intensity proportional to the increasing UV dose. The quantitation for this dot blot is shown in Figure 6.5, where the amount of damage (expressed as IOD) induced by UV was plotted for each DNA concentration used. These two DNA concentrations were chosen as starting points for optimizing the amount of photolyase needed to produce visible repair on the membrane. The *E. coli* photolyase was used for the optimization experiments. The UV irradiated DNA solution was blotted unto the membranes and prepared as described in Section 6.3.3. The membranes were either soaked in 2ml of a diluted stock of pure *E. coli* photolyase at 1ng/µl, or 2ml lysate produced from a 3-day grown culture of the homologously expressed Phr2 strain as described in Chapter 3. Control blots were incubated with water only. The blots were either illuminated with a white light source or incubated in the dark for 2 hr, followed by rinsing the membranes in TBS-T to remove any residual photolyase. The anti-thymine dimer antibody reacted

specifically with CPDs and not (6-4) photoproducts; therefore, the decrease in intensity of the dots was an indication of repair of CPDs only. The developed blots are as shown in Figure 6.6.

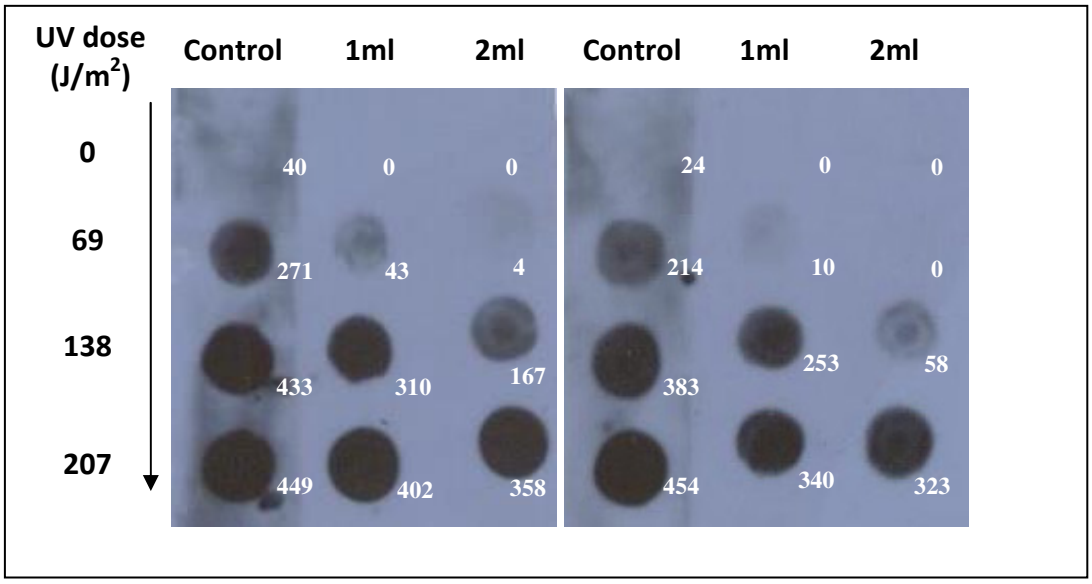


Figure 6.6: Dot blot showing repair of dimers using the *E. coli* photolyase (left) and the Phr2 overexpression strain (right). For *E. coli,* 1 and 2ml solutions containing 1ng/µl of photolyase was used. 1 and 2ml of whole cells of the Phr2 overexpression strain were used. Numbers inset represent IOD values for each dot.

The finished blots as shown in Figure 6.6 demonstrate that both the *E. coli* and *Hfx. volcanii* Phr2 photolyase were capable of CPD repair as indicated by the decreased intensity of the incubated blots compared to the control blots. Repair was efficient up to a UV dose of 138 J/m^2, when the decrease in intensity was more pronounced for the Phr2 expression strain. This provided a first indication that the Phr2 expression strain may be more efficient in removing UV-induced DNA dimers. The DNA concentration was adjusted to 0.025mg/ml for the rest of the immunoassay experiments.

Following the successful demonstration that *E. coli* photolyase and the homologously expressed Phr2 strain were both capable of repairing UV-induced DNA damages using this immunoassay, the re-folded Phr2 protein, as described in Chapter 5, was used in this experiment. The blots were also incubated in the dark to assess differences between light and dark repair. Figure 6.7 shows a graphical comparison of IOD vs. UV exposure time between *E. coli* photolyase, the re-folded

Phr2 protein and the Phr2 expression strain, where 3 replicates were performed for each. As shown, the control (water only) strain showed an increasing level of intensity as UV exposure increased; this served as a baseline of comparison for the amount of DNA damage at each UV exposure time. The trend for the re-folded protein did not differ from that of the control, whereas the trend for *E. coli* photolyase differed as the IOD values were consistently lower at a UV dose of 69 J/m^2. The trend for the Phr2 expression strain differed to a greater extent than *E. coli* photolyase, as the IOD values remained the lowest throughout the experiment. The re-folded Phr2 protein did not appear to repair DNA dimers as the increase in IOD remained very high from 0 to 207 J/m^2, whereas repair was observed for both the *E. coli* photolyase and the Phr2 expression strain. A two-way ANOVA test indicated that the source of variation from UV exposure was considered to be an extremely significant factor and accounted for a large proportion of the total variation. The interaction factor which showed how significant the interaction was between group (*E. coli* photolyase, re-folded Phr2, Phr2 expression, and control) and UV exposure, was also found to be significant; this was due to the increasing differences in IOD between the Phr2 expression strain and the other groups with an increasing UV exposure. A post test was use to compare each pair of groups for each UV exposure time, indicating that the means of ±SEM for the Phr2 strain were significantly different compared to the control and re-folded Phr2 UV doses of 138 and 207 J/m^2 ($P < 0.001$), and the *E. coli* photolyase at 207 J/m^2 ($P < 0.01$). This suggested that while the *E. coli* photolyase also removed dimers, it was not as efficient as the Phr2 expression strain. No significant differences were found between the means of the control, the re-folded Phr2 and the *E. coli* photolyase. A visual inspection of the developed blots shown can be seen in Figure 6.8, showing the differences in the intensity of the dots for each strain at various UV doses. The dots were fainter for the Phr2 expression strain with light exposure (lanes 3-6), implicating the lower amount of dimers present.

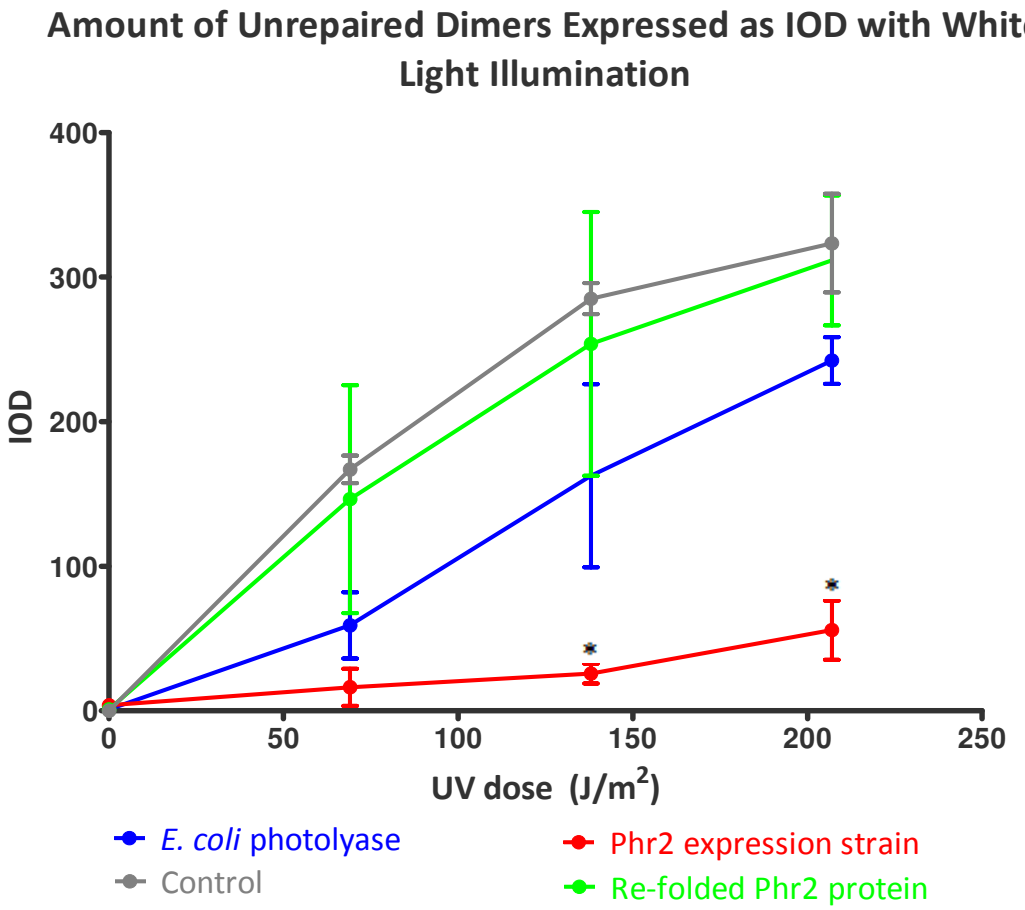
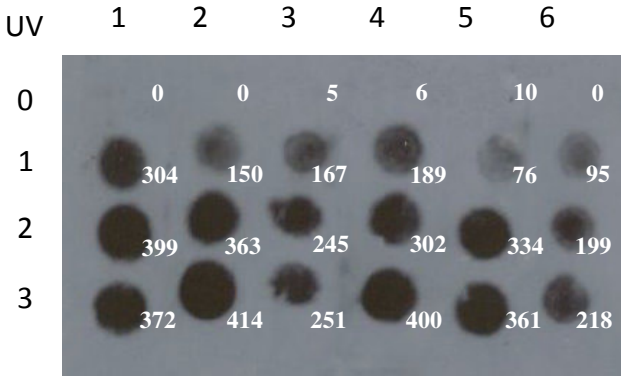


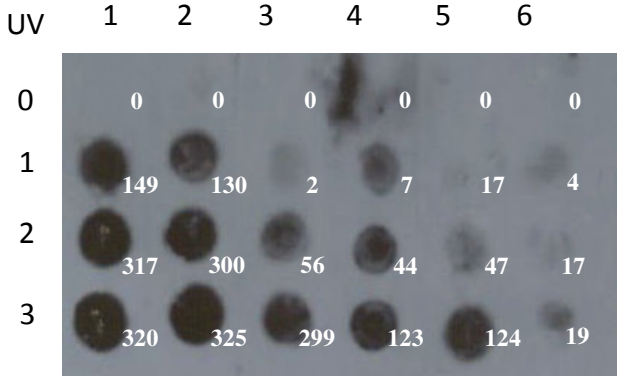
Figure 6.7: **Comparison of the repair of thymine dimers with time for *E. coli* photolyase, Phr2 expression strain and the re-folded Phr2 protein.** Control strain was incubated with water only. All IODs were normalized against the background value at t = 0 min. UV irradiated herring sperm DNA was blotted onto membrane (UV output 1.15 J/m^2/s), followed by incubation with mouse monoclonal antibody raised against thymine dimers, diluted 1:500 in 5% milk/TBS-T, for 1 hr. Following a 4 x 10min wash in TBS-T buffer, membrane was incubated with a peroxidase-conjugated anti-mouse antibody, diluted 1:4000, for 1 hr, washed 4 x 10min in TBS-T before overlaying the chemiluminescent HRP substrate on top, DNA side up, for 5 min. To evaluate antibody binding, the blots were transferred to an x-ray film and developed, followed by using LabWorks to quantitate damage. Values are presented as means ± SEM from 4 independent groups, each containing 3 replicates. *Statistical significance at $P < 0.001$ (compared to the control and the re-folded Phr2).

***E. coli* Photolyase**



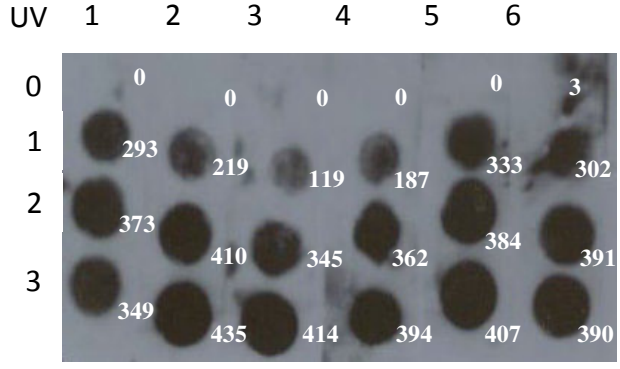
1 – light water control
2 – dark water control
3 – dark (40ng)
4 – dark (100ng)
5 – light (40ng)
6 – light (100ng)

Phr2 Expression Strain



1 – light water control
2 – dark water control
3 – dark (40ng)
4 – dark (100ng)
5 – light (40ng)
6 – light (100ng)

Re-folded Phr2 Protein



1 – light water control
2 – dark water control
3 – dark (40ng)
4 – dark (100ng)
5 – light (40ng)
6 – light (100ng)

Figure 6.8: Representative dot blots showing the detection of thymine dimers with time. UV irradiated herring sperm DNA was blotted onto membrane (UV output 1.15 J/m^2/s), followed by incubation with mouse monoclonal antibody raised against thymine dimers, diluted 1:500 in 5% milk/TBS-T, for 1 hr. Following a 4 x 10min wash in TBS-T buffer, membrane was incubated with a peroxidase-conjugated anti-mouse antibody, diluted 1:4000, for 1 hr, washed 4 x 10min in TBS-T before overlaying the chemiluminescent HRP substrate on top, DNA side up, for 5 min. To evaluate antibody binding, the blots were transferred to an x-ray film and developed, followed by using LabWorks to quantitate damage. Numbers inset represent IOD values for each dot.

Figure 6.9 shows the results obtained when 3 expression strains were used: Phr2, Phr1 and E3. IOD values for the Phr2 strain were again low compared to the control strain from at UV doses from 69 to 138 J/m^2. The IOD values for the Phr1 strain were consistently high and comparable to the control values. The extent of repair from the E3 strain was not as pronounced as the Phr2 strain, which showed obvious superiority in the removal of dimers as indicated by the lowest IOD values obtained. As this experiment was only performed once, no statistical analysis was carried out.

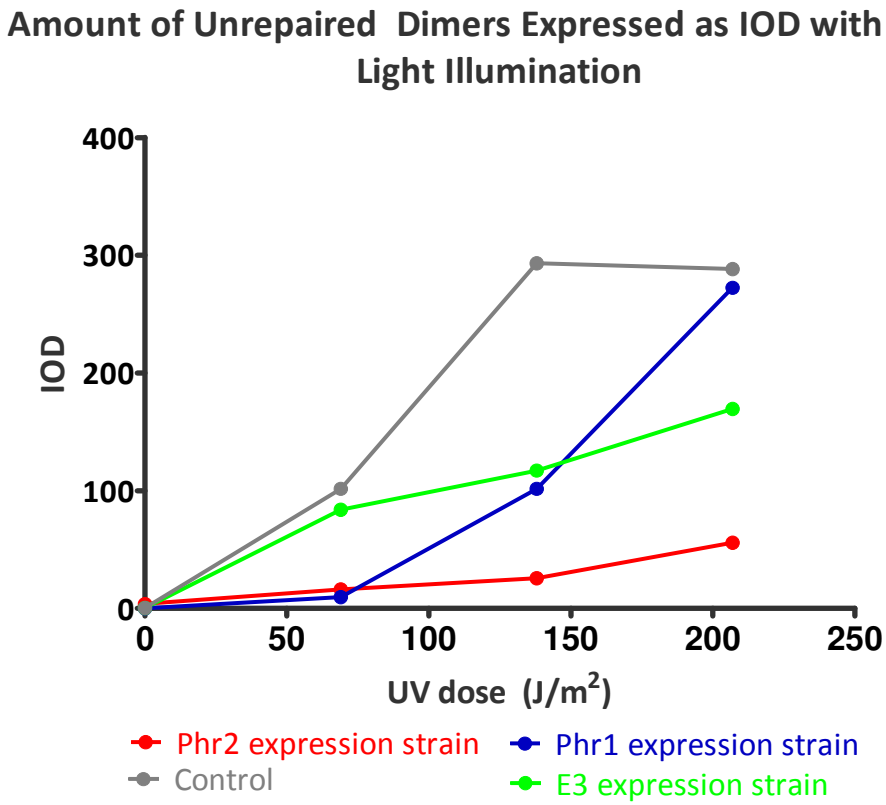


Figure 6.9: Comparison of repair of thymine dimers with time for the Phr2, Phr1 and E3 expression strains. Control strain was incubated with water only. All IODs were normalized against the background value at t = 0 min. UV irradiated herring sperm DNA was blotted onto membrane (UV output at 1.15 J/m^2/s), followed by incubation with mouse monoclonal antibody raised against thymine dimers, diluted 1:500 in 5% milk/TBS-T, for 1 hr. Following a 4 x 10min wash in TBS-T buffer, membrane was incubated with a peroxidase-conjugated anti-mouse antibody, diluted 1:4000, for 1 hr, washed 4 x 10min in TBS-T before overlaying the chemiluminescent HRP substrate on top, DNA side up, for 5 min. The chemiluminescent HRP substrate was used as the detection agent. To evaluate antibody binding, the blots were transferred to an x-ray film and developed, followed by using LabWorks to quantitate damage.

6.5 Discussion

Numerous *in vitro* assays are available for the detection of photolyase activity from purified enzymes. One of the simpler methods is the change in absorption at 265nm between intact DNA and dimerized thymine residues (Kato et al. 1997) as demonstrated for the purified *T. thermophilus* (Fujihashi et al. 2007) and the *S. tokodaii* photolyases (Kato et al. 1997). However, the sensitivity of this assay is directly correlated with the purity of the enzyme, making it unsuitable for measuring activity of the *Hfx. volcanii*, Phr2 protein. Another *in vitro* method is by the use restriction enzymes assay by which the fraction of molecules of DNA cleaved by endonucleases is an indication of the photoreactivating activity of photolyase (Dutta et al. 1993). This method has been modified for use with the Trevigen® protocol for assessing the efficiency of *E. coli* photolyases in repairing supercoiled plasmid DNA irradiated with UV (http://www.trevigen.com/dnadamage/ecoliphotolyase.php) . In this protocol, the repair is measured as a loss in conversion of the supercoiled plasmids to relaxed, open circle form post treatment with a CPD-specific endonuclease T4-PDG. However, the production of a pure enzyme is again crucial as contaminanting nucleases can affect the results. Other assays described, such as the high-performance liquid chromatography methods used for *E. coli* photolyase (Nakayama et al. 2004) again requires the use of a pure, soluble protein. In order to carry out functional analysis of the *Hfx. volcanii* Phr2 protein, assays were chosen for reproducibility without the requirement for a pure protein.

In Chapter 3, results indicated that while the homologous protein expression of Phr2 was low, the transformation of the construct carrying the cloned *phr2* gene into the H98 strain had been successful. For this reason, an *in vivo* assay was developed to look for the demonstration that the strain carrying the *phr2* gene was more UV resistant following exposure to white light, hence providing evidence that *phr2* encodes a photolyase. The UV survival assay was optimized to allow for ease of use and reproducibility. Results indicated that the Phr2 expression strain, upon exposure to white light, was more resistant to UV irradiation compared to either the Phr1 or the E3 expression strains. When the 3 strains were incubated with white light, there was an enhancement of survival for the Phr2 expression strain up to a UV dose of 207 J/m^2, whereas under the same conditions the Phr1 or the E3 expression strains

were only resistant up to 69 J/m^2. This strongly demonstrated that photorepair was a crucial factor in the strain expressing the Phr2 protein. In the dark, similar levels of repair was observed for all the strains, suggesting a less efficient dark repair system.

The superiority of survival for the Phr2 strain illuminated with white light was observed with repeated trials, giving these qualitative data credibility and ease of replication. These data were complemented by the cell viability experiments using an ATP assay, which allowed for quantitative analysis of survival. The Phr2 expression strain, exposed to white light following UV exposure, resulted in higher % ATP levels, thus correlating to an enhancement of survival compared to the Phr2 expression strain not illuminated with white light. White light enhancement of survival was not observed for the E3 expression strain. The Phr2 and E3 expression strains experienced decreases in cell numbers when incubated in the dark throughout the experiment. % ATP for the Phr2 expression strain exposed to white light decreased after a UV dose of 276 J/m^2, showing some consistency with the UV survival assay where colony formation ceased after a UV dose of 207 J/m^2. Due to time constraints, the ATP assay for the Phr1 expression strain was not conducted.

To account for the fact that % ATP levels were sometimes above 100% for this assay, factors such as environmental stresses may be taken into consideration. Any changes in media used or temperature may have impeded cell growth prior to the commencement of the experiment. But more importantly, white light illumination may have offered a form of stress relief for the cells undergoing changes in their environment. The presence of over-expressed Phr2 proteins may have contributed to this protective mechanism induced by white light. To further investigate this link between stress-protection and the Phr2 protein, growth curves experiments may be conducted to determine whether the presence of white light and Phr2 together causes higher cell growth under various environmental stresses.

Results from the heterologous expression of Phr2 in Chapter 5 demonstrated the difficulty in obtaining a soluble protein. The insolubility of Phr2 meant that re-folding treatments were required, with no guarantee that the protein would re-gain the correct structure as protein misfolding is a common phenomenon. CD

spectroscopy showed that the re-folded Phr2 was highly helical, a feature shared amongst many members of the photolyase/crytochrome superfamily. The use of the immunoassay offered means of testing the activity of the re-folded protein. The dot blot immunoassay, adapted from that of McCready, measured repair of DNA photoproducts, and was sensitive enough to measure repair of damage induced by low doses of UV-C. It has been used to measure repair in human cells, yeasts and a variety of other organisms (McCready 1999). The DNA used for the assay does not have to be intact and is easy to prepare; this is different from fluorescence methods where high integrity of DNA is required (Kundu et al. 2002) and complicated and time-consuming preparation of the dimer substrates is often necessary.

The assay was first optimized with the commercially prepared *E. coli* photolyase, which provided positive findings that light repair of DNA dimers was more efficient than dark repair, providing evidence that the assay was working properly. The assay continued with the re-folded Phr2 protein to establish its activity, therefore, its correctly folded form. Results demonstrated that the re-folded protein had no function in photorepair, which was most likely due to misfolding during renaturation where the protein did not have the appropriate residues in the right positions to bind the FAD molecule. A second co-factor was also absent, as no respective spectrum was observed for either MTHF or 8-HDF. This was not surprising, however, as homology modeling of photolyase from *Hfx. volcanii* has predicted for the presence of a 8-HDF molecule which *E. coli* does not produce. Nevertheless, confident and convincing results were gathered for the Phr2 expression strain, where repair was superior compared to the *E. coli* photolyse, Phr1 and the E3 expression strains when white light illumination was present. The Phr2 expression strain produced the lowest IOD values, translating to the most efficient repair of DNA dimers. The IOD values for the Phr1 expression were three fold higher than those of the Phr2 expression strain when the UV dose was increased beyond 138 J/m^2. This correlates well to the light repair findings from the UV survival assay, where colony formation for the Phr1 expression strain ceased after a UV dose of 69 J/m^2.

Due to the way that the assay was conducted, it was not possible to construct survival curves for each strain. As a refinement to the immunoassay, the UV-

irradiated blots containing the damaged DNA can be illuminated with white light for an extended amount of time to look at the time-dependant activity of photolyase. The % lesions remaining can then be quantified over different periods of white light exposure time post UV irradiation and provide a baseline value (white light exposure at time 0) for comparison.

Overall, all three assays produced convincing results that complemented each other very well. They firstly demonstrated the general trend that white light illumination enhanced survival of UV irradiated cells. More importantly, the assays confirmed that the over-expression of the Phr2 protein led to improved survival of cells in the presence of white light. This correlates well with previous studies where it was shown that DNA repair was compromised by knocking out the *phr2* gene. From these assays, it was also demonstrated that the over-expression of Phr1 did not produce the same level of enhancement in strain survival. These results clearly demonstrate that the *phr2* gene encodes a functional photolyase, a function lacking for the *phr1* gene product.

General Discussion

The ultimate aim of this PhD project was the identification and functional characterisation of novel photolyase-like genes in the model halophilic archaeon, *Hfx. volcanii*. This study conclusively demonstrates that the novel *phr2* gene identified in *Hfx. volcanii* encodes a functional CPD photolyase, whereas the *phr1* gene is not involved in CPD photorepair. Attempts at obtaining sufficiently large and pure levels of the protein products of these two genes proved unsuccessful, preventing the structural analyses of halophilic photolyases. All the approaches used in this project were directed at creating a logical and sensible flow of experiments to determine the novel genes responsible for photoreactivation in *Hfx. volcanii*.

7.1 Bioinformatics and structure modelling

The overall objective of this study was to obtain initial evidence for the presence of a gene encoding photolyase in *Hfx volcanii*, and to gather preliminary sequence and structural data on this putative enzyme. The putative gene homologs, *phr1* and *phr2* were identified from the then unfinished and un-annotated genome of *Hfx. volcanii* using the known sequences from *Halobacterium* sp. NRC-1. However, the contig arrangement of the genomic sequences resulted in the distribution of the *phr2* gene sequence on two separate contigs with a 15 base pair overlap. These putative transcripts were translated to protein sequences. Based on sequence alignments to representative members of the photolyases/cryptochrome superfamily from all three domains, a high degree of sequence similarity at the C-terminal region of about 150 amino acids was shown. This is consistent with previous findings of a well conserved FAD domain towards the C-terminal region in all photolyases and cryptochromes analyzed to date (Sancar 2003). In terms of sequence identity, *Hfx. volcanii* Phr2 was most closely related to Phr2 and Phr1 from *Halobacterium* sp. NRC-1; but this was not surprising as they are both species of the *Halobacteriaceae* family, where one would expect to find close similarities in their genomes. Compared to the corresponding sequences from *E. coli* and *A. nidulans*, *Hfx volcanii* Phr2 resembled a deazaflavin type photolyase where 8-HDF occupies the light-harvesting chromophore position. Homology modelling, based on the structure of *A. nidulans* photolyase, was used to create a hypothetical photolyase structure from

Hfx. volcanii. A close examination of the interacting residues with the light harvesting chromophore revealed the major conserved residues in *Hfx. volcanii* that bound 8-HDF, again suggesting the presence of a deavaflavin type photolyase for *Hfx. volcanii*. Meanwhile, *Hfx. volcanii* Phr1 appears to resemble the ssDNAphotolyases, which were previously believed to be crytochromes based on sequence similarity to these blue-light photoreceptors.

7.2 Homologous expression and purification of Phr proteins

The homologous expression of halophilic proteins was attempted in their native environment to help protein folding into their correct and functional conformations as well as undergoing any post-translational modifications. However, limited results were obtained using this system, which could be ascribed to several factors outlined below.

In this study, the polymerase chain reaction (PCR) was used to amplify the *phr* genes, followed by the cloning of these genes into intermediate vectors, pGEM-T and pIL11, where the latter was used to recover the *fdx* promoter prior to the cloning of the *fdx-phr* constructs into the *E coli/Hfx. volcanii* shuttle vector, pTA233. The de-methylated constructs, generated by the transformation into *dam-/dcm- E. coli* strain, were finally transformed into the native *Haloferax* H98 strain for expression. The expression yield was low, most likely due to the low copy number of the shuttle vector used in this experiment.

The lack of inducible promoters and suitable expression vectors, coupled with an incomplete knowledge of the intracellular biochemistry and the genetics of these organisms reduced the chance of success in obtaining high expression levels of the target proteins. With the absence of an efficient system for the expression of halophilic proteins, downstream applications such as purification and functional characterization were severely hindered. A strategy was employed to obtain as much soluble protein as possible by modification of the shuttle vector to include a polyhistidine tag for affinity purification. Numerous methods were tried and tested, eventually leading to one that produced a soluble protein of the right size and of high

purity as visualized on SDS-PAGE. Mass spectroscopy analysis, however, did not correlate with the expected finding, as the identity of the bound protein was not Phr2 but rather a native protein shown to contain a long string of histidine residues within its sequence. This suggested that the level of Phr2 expression was extremely low and not in sufficient quantity to be purified. In hindsight, this may not have been a surprise finding, as previous studies on photolyase from yeast and algae have indicated low levels of expression *in vivo*, leading to unsuccessful purification of the proteins for functional or structural analyses (Beukers et al. 2008). This is also true in *E. coli* and yeast, where the number of photolyase molecules is found to be between 20-30, and 75-100, respectively (Harm et al. 1968; Yasui and Laskowski 1975). This is, however, compensated by the extreme efficiency of these enzymes for the removal of DNA damage. This is demonstrated by the improved *in vivo* photorepair efficiency of the homologous Phr2 overexpression strain in all three assays used.

7.3 Heterologous expression and re-folding

The advantage of a heterologous expression system for any protein is largely due to the extremely well characterized bacterial expression systems utilizing *E. coli* as a host strain. The rapid growth, high cell densities, well-known and developed genetics and transformation techniques, and the availability of a large number of commercial expression vectors with inducible promoters that offer tightly controlled expression to high levels are some of the benefits of such system. With this in mind, heterologous expression of the *phr* genes was conducted utilizing the widely popular pET system. Previous studies within the lab have demonstrated high yields of heterlogously expressed proteins compared to expression in the native halophilic organism (Jolley et al. 1996; Connaris et al. 1999). The problems of expressing halophilic proteins in a mesophilic environment include the incorrect post-translational changes and misfolding of the protein resulting in an inactive enzyme. Most likely, however, is the formation of insoluble aggregates that require either reactivation or re-folding. This could create a large amount of work in identifying the correct conditions under which the protein will fold properly.

In this current study, the Phr proteins were heterologously expressed in *E. coli* using the pET system, whereby the *phr* genes were cloned into the high copy number vector with an inducible T7 promoter. Despite numerous changes in growth conditions and temperatures, concentrations of inducer, and host strains, the highly expressed proteins remained as insoluble aggregates. To obtain a more homogenous population of proteins, affinity purification was performed under denaturing conditions. The purified fraction was solubilised in 8M urea followed by treatments with various re-folding buffers, eventually yielding a Phr2 protein of high purity as seen in SDS-PAGE analyses. Circular dichroism spectroscopy revealed the presence of secondary structures with a high ratio of α-helixes. The protein, however, did not appear to bind FAD as indicated by the absence of the expected absorption spectrum of flavin.

7.5 Functional analysis

This final study is of great significance because it demonstrates conclusively that the novel *phr2* gene in *Hfx. volcanii* encodes a functional CPD photolyase, whereas the *phr1* gene is not involved in CPD photorepair. The lack of pure soluble proteins indirectly affected and, possibly, limited the types of assays that could be performed. However, it also allowed for a more basic approach to assessing activity while circumventing those that were deemed incompatible. An *in vivo* UV survival assay was developed to assess the photosensitivity of the homologously expressed strains carrying the Phr2 protein upon exposure to UV followed by white light. This assay produced clear and visual qualitative evidence of photorepair, where the Phr2 overexpression strain was shown to be more UV resistant in the presence of white light. The data were complemented by a cell viability assay to quantify the number of surviving cells post UV irradiation using ATP release as an endpoint. This assay strongly demonstrated that photoreactivation-coupled repair played a bigger role in survival compared to dark repair in the Phr2 overexpression strain as indicated by the greater amount of ATP released by live cells. For the *in vitro* analysis of protein function, an antibody-based assay was opitmized for use with the Phr2, Phr1 and E3 overexpression strains as well as the re-folded Phr2 protein. Results demonstrated that, unlike the overexpression strains, the re-folded protein had no activity in photorepair, which was most likely due to misfolding during renaturation preventing

binding to the FAD molecule. As *E. coli* does not synthesize 8-HDF, it also meant that the photoantenna chromophore was missing, thereby reducing the efficiency of photoreactivation. Nevertheless, results were compelling for the Phr2 expression strain, where repair was superior compared to both the Phr1 and the E3 expression strains when white light illumination was present. Overall, the results gathered from the three assays complemented one another, clearly demonstrating that the *phr2* gene encoded a functional photolyase, a function that was lacking for the *phr1* gene product.

7.6 Future work

Whilst the goal of achieving evidence for the function of the encoded Phr2 protein has been demonstrated, there remains obvious extensions and improvements to this Ph.D. project. For one, the attainment of the crystal structure of photolyase from *Hfx. volcanii* would be the first of its kind, providing new revelation to what is currently known of these enzymes from halophilic archaea. It is probable that the different features of halophilic photolyases may provide answers to the extreme efficiency of these enzymes compared to mesophilic ones. On a broader scale, it would be useful to analyze the structural data in parallel with the current phylogenomics of the photolyase/cryptochrome superfamily to determine the evolutionary history of these enzymes within the archaeal domain.

Obtaining a crystal structure requires the production of a soluble, functional protein. This was not achieved in this project due to numerous reasons, including an insufficient levels of homologous expression of halophilic proteins and the high likelihood of producing inactive and insoluble proteins in mesophilic hosts. With more time, it would have been worthwhile to try a different tag system for the purification of the homologously expressed protein. Isoelectric point focusing can be applied afterwards to separate different species within a sample to allow for the purification of the target protein, provided that the pI of these proteins are different enough to do so. In addition to the tag strategy, the tightly controlled *tna* promoter may prove to be useful for the over-expression of proteins that often result in the formation of inclusion bodies (Large et al. 2007). This can be done by growing cells

to a high density in media without tryptophan, followed by using induction with tryptophan late in exponential phase to allow higher production of the target protein (Large et al. 2007). With the heterologous expression system, many components can be further fine-tuned to produce soluble proteins, such as changes in host strains and vectors used. Once a soluble protein is achieved, various assays can then be applied to elucidate the function, such as the one applied to *S. tokodaii* looking at the difference in absorption spectra of the repaired and unrepaired DNA. In addition, spectroelectrochemical titration studies can be applied to the protein to determine the redox transitions of FAD, thus providing further evidence that the encoded protein is a functional photolyase containing the reduced FAD in its active form and not an oxidized FAD like the structurally similar crytochromes (Balland et al. 2009).

Other areas of expansion would include the creation of deletion mutants in *phr2* and *phr1* in *Hfx. volcanii* and observe the differences in survival of wild type and mutant strains to complement previous data gathered from studies with *Halobacterium* sp. NRC-1. From a health-related aspect, the expression of an archaeal photolyase in a mammalian system carries with it an implication for human skin health, as placental mammals have been shown not to contain this enzyme. The transgenic mouse has proved to be a useful model for the analysis of the relative contribution of CPDs and (6-4) photoproducts to the detrimental effects of UV light. These mice were generated to ubiquitously express CPD photolyase, (6-4) photolyase, or both, to assess light-dependent repair of the different photoproducts in the skin. Results revealed an enhancement of repair of photoproducts in the dermis and epidermis for the transgenic mice, leading to superior survival when treated with visible light post UV irradiation. The CPD photolyse-expressing mice also exhibited better survival than those expressing (6-4) photolyase; this further adds to the finding that CPDs are the principal cause of nonmelanoma cancer and that photolyases can be used as very effective tools in combating skin cancer (Schul et al. 2002; Jans et al. 2005).

To conclude, further investigation on the optimization of the expression procedure of the Phr proteins is required, as this would allow for several downstream applications as discussed above, but most importantly, for the structural elucidation of the first halophilic photolyase.

References

Ahmad, M. and Cashmore, A. R. (1993) *Hy4* Gene of *A. thaliana* Encodes a Protein with Characteristics of a Blue-Light Photoreceptor. *Nature.*, **366**: 162-166.

Ahmad, M., Jarillo, J. A., Klimczak, L. J., Landry, L. G., Peng, T., Last, R. L. and Cashmore, A. R. (1997) An enzyme similar to animal type II photolyases mediates photoreactivation in *Arabidopsis*. *Plant Cell.*, **9**: 199-207.

Aivaliotis, M., et al. (2007) Large-scale identification of N-terminal peptides in the halophilic archaea *Halobacterium salinarum* and *Natronomonas pharaonis*. *J. Proteome. Res.*, **6**: 2195-204.

Allers, T. and Mevarech, M. (2005) Archaeal genetics - the third way. *Nat. Rev. Genet.*, **6**: 58-73.

Allers, T., Ngo, H. P., Mevarech, M. and Lloyd, R. G. (2004) Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.*, **70**: 943-953.

AL-Mailem, D. (2006) *The 2-oxoacid Dehydrogenase Complex of Hfx. volcanii*. Ph.D.thesis, University of Bath, Bath.

Bab-Dinitz, E., Shmuely, H., Maupin-Furlow, J., Eichler, J. and Shaanan, B. (2006) *Haloferax volcanii* PitA: an example of functional interaction between the Pfam chlorite dismutase and antibiotic biosynthesis monooxygenase families? *Bioinformatics.*, **22**: 671-675.

Baliga, N. S., Bjork, S. J., Bonneau, R., Pan, M., Iloanusi, C., Kottemann, M. C., Hood, L. and DiRuggiero, J. (2004) Systems level insights into the stress response to UV radiation in the halophilic archaeon *Halobacterium* NRC-1. *Genome. Res.*, **14**: 1025-1035.

Baliga, N. S., et al. (2004) Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome. Res.*, **14**: 2221-2234.

Baliga, N. S., et al. (2002) Coordinate regulation of energy transduction modules in *Halobacterium* sp. analyzed by a global systems approach. *Proc. Natl. Acad. Sci. U. S. A.*, **99**: 14913-14918.

Balland, V., Byrdin, M., Eker, A. P. M., Ahmad, M. and Brettel, K. (2009) What makes the difference between a cryptochrome and DNA photolyase? A spectroelectrochemical comparison of the flavin redox transitions. *J. Am. Chem. Soc*. **131**: 426-427.

Barns, S. M., Delwiche, C. F., Palmer, J. D. and Pace, N. R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl. Acad. Sci. U. S. A.*, **93**: 9188-9193.

Bateman, A., et al. (2004) The Pfam protein families database. *Nucleic. Acids. Res.*, **32**: D138-141.

Batschauer, A. (1993) A Plant Gene for Photolyase - an Enzyme Catalyzing the Repair of Uv-Light-Induced DNA-Damage. *Plant. J*. **4**: 705-709.

Benachenhou, N. and Baldacci, G. (1991) The gene for a halophilic glutamate dehydrogenase: sequence, transcription analysis and phylogenetic implications. *Mol. Gen. Genet.*, **230**: 345-352.

Berquist, B. R. and DasSarma, S. (2003) An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, *Halobacterium* sp. strain NRC-1. *J. Bacteriol.*, **185**: 5959-5966.

Beukers, R., Eker, A. P. and Lohman, P. H. (2008) 50 years thymine dimer. *DNA Repair (Amst).* **7**: 530-543.

Bidle, K. A., Kirkland, P. A., Nannen, J. L. and Maupin-Furlow, J. A. (2008) Proteomic analysis of *Haloferax volcanii* reveals salinity-mediated regulation of the stress response protein PspA. *Microbiology.*, **154**: 1436-1443.

Bieger, B., Essen, L. O. and Oesterhelt, D. (2003) Crystal structure of halophilic dodecin: a novel, dodecameric flavin binding protein from *Halobacterium salinarum*. *Structure.*, **11**: 375-385.

Bischoff, K. M. and Rodwell, V. W. (1997) 3-Hydroxy-3-methylglutaryl-coenzyme A reductase of *Haloferax volcanii*: role of histidine 398 and attenuation of activity by introduction of negative charge at position 404. *Protein. Sci.*, **6**: 156-161.

Bisle, B., et al. (2006) Quantitative profiling of the membrane proteome in a halophilic archaeon. *Mol. Cell. Proteomics.*, **5**: 1543-1558.

Bitan-Banin, G., Ortenberg, R. and Mevarech, M. (2003) Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. *J. Bacteriol.*, **185**: 772-778.

Blecher, O., Goldman, S. and Mevarech, M. (1993) High expression in *Escherichia coli* of the gene coding for dihydrofolate reductase of the extremely halophilic archaebacterium *Haloferax volcanii*. Reconstitution of the active enzyme and mutation studies. *Eur. J. Biochem.*, **216**: 199-203.

Bohm, G. and Jaenicke, R. (1994a) Relevance of sequence statistics for the properties of extremophilic proteins. *Int. J. Pept. Protein. Res.*, **43**: 97-106.

Bohm, G. and Jaenicke, R. (1994b) A structure-based model for the halophilic adaptation of dihydrofolate reductase from *Halobacterium volcanii*. *Protein. Eng.*, **7**: 213-220.

Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodriguez-Valera, F., Pfeiffer, F. and Oesterhelt, D. (2006) The genome of the square archaeon *Haloquadratum walsbyi* : life at the limits of water activity. *BMC. Genomics.*, **7**: 169.

Bolhuis, H., Poele, E. M. and Rodriguez-Valera, F. (2004) Isolation and cultivation of Walsby's square archaeon. *Environ. Microbiol.*, **6**: 1287-1291.

Bonnete, F., Madern, D. and Zaccai., G. (1993) A biophysical structure and solvent interactions in native and recombinant enzyme. *J. Chem. Soc., Faraday Trans.*, **89**: 2659–2666.

Boubriak, I., Ng, W. L., Dassarma, P., Dassarma, S., Crowley, D. J. and McCready, S. J. (2008) Transcriptional responses to biologically relevant doses of UV-B radiation in the model archaeon, *Halobacterium* sp. NRC-1. Saline Systems. 4: 13.

Bouyoub, A., Barbier, G., Querellou, J. and Forterre, P. (1995) A putative SOS repair gene (dinF-like) in a hyperthermophilic archaeon. *Gene.*, **167**: 147-149.

Bradford, M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, **7**: 248-254.

Brautigam, C. A., Smith, B. S., Ma, Z., Palnitkar, M., Tomchick, D. R., Machius, M. and Deisenhofer, J. (2004) Structure of the photolyase-like domain of cryptochrome 1 from *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.*, **101**: 12142-12147.

Brochier, C., Bapteste, E., Moreira, D. and Philippe, H. (2002) Eubacterial phylogeny based on translational apparatus proteins. *Trends. Genet.*, **18**: 1-5.

Brochier, C., Philippe, H. and Moreira, D. (2000) The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome. *Trends. Genet.*, **16**: 529-533.

Brochier-Armanet, C., Boussau, B., Gribaldo, S. and Forterre, P. (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the *Thaumarchaeota*. *Nat. Rev. Microbiol.*, **6**: 245-252.

Brody, J. R. and Kern, S. E. (2004) Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. *Biotechniques.*, **36**: 214-216.

Brown, J. R. and Doolittle, W. F. (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc. Natl. Acad. Sci. U. S. A.*, **92**: 2441-2445.

Brown, J. R., Robb, F. T., Weiss, R. and Doolittle, W. F. (1997) Evidence for the early divergence of tryptophanyl- and tyrosyl-tRNA synthetases. *J. Mol. Evol.*, **45**: 9-16.

Brown, R. E., Jarvis, K. L. and Hyland, K. J. (1989) Protein measurement using bicinchoninic acid: elimination of interfering substances. *Anal. Biochem.*, **180**: 136-139.

Brudler, R., et al. (2003) Identification of a new cryptochrome class. Structure, function, and evolution. *Mol. Cell.* . **11**: 59-67.

Burns, D. G., Camakaris, H. M., Janssen, P. H. and Dyall-Smith, M. L. (2004) Cultivation of Walsby's square haloarchaeon. *FEMS. Microbiol. Lett.*, **238**: 469-473.

Burns, D. G., Janssen, P. H., Itoh, T., Kamekura, M., Li, Z., Jensen, G., Rodriguez-Valera, F., Bolhuis, H. and Dyall-Smith, M. L. (2007) *Haloquadratum walsbyi* gen. nov., sp. nov., the square haloarchaeon of Walsby, isolated from saltern crystallizers in Australia and Spain. *Int. J. Syst. Evol. Microbiol.*, **57**: 387-392.

Cabrera, J. A., Bolds, J., Shields, P. E., Havel, C. M. and Watson, J. A. (1986) Isoprenoid synthesis in *Halobacterium halobium*. Modulation of 3-hydroxy-3-methylglutaryl coenzyme a concentration in response to mevalonate availability. *J. Biol. Chem.*, **261**: 3578-3583.

Camacho, M., Rodriguez-Arnedo, A. and Bonete, M. J. (2002) NADP-dependent isocitrate dehydrogenase from the halophilic archaeon *Haloferax volcanii*: cloning, sequence determination and overexpression in *Escherichia coli*. *FEMS. Microbiol. Lett.*, **209**: 155-160.

Caton, T. M., Caton, I. R., Witte, L. R. and Schneegurt, M. A. (2009) Archaeal diversity at the great Salt Plains of Oklahoma described by cultivation and molecular analyses. *Microbial Ecology*.[Epub ahead of print].

Cavalier-Smith, T. (2002) The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.*, **52**: 7-76.

Cendrin, F., Chroboczek, J., Zaccai, G., Eisenberg, H. and Mevarech, M. (1993) Cloning, sequencing, and expression in *Escherichia coli* of the gene coding for malate dehydrogenase of the extremely halophilic archaebacterium *Haloarcula marismortui*. *Biochemistry.*, **32**: 4308-4313.

Chanderkar, L. P. and Jorns, M. S. (1991) Effect of flavin structure and redox state on catalysis by and flavin-pterin energy transfer in *Escherichia coli* DNA photolyase. *Biochemistry.*, **30**: 745-754.

Charlebois, R. L., Lam, W. L., Cline, S. W. and Doolittle, W. F. (1987) Characterization of Phv2 from *Halobacterium volcanii* and Its Use in Demonstrating Transformation of an Archaebacterium. *Proc. Natl. Acad. Sci. U. S. A.*, **84**: 8530-8534.

Charlebois, R. L., Schalkwyk, L. C., Hofman, J. D. and Doolittle, W. F. (1991) Detailed Physical Map and Set of Overlapping Clones Covering the Genome of the Archaebacterium *Haloferax volcanii* Ds2. *J. Mol. Biol.*, **222**: 509-524.

Chen, A. and Poulter, C. D. (1993) Purification and characterization of farnesyl diphosphate/geranylgeranyl diphosphate synthase. A thermostable bifunctional enzyme from *Methanobacterium thermoautotrophicum*. *J. Biol. Chem.*, **268**: 11002-11007.

Cheng, L., Qiao, D. R., Lu, X. Y., Xiong, Y., Bai, L. H., Xu, H., Yang, Y. and Cao, Y. (2007) Identification and expression of the gene product encoding a CPD photolyase from *Dunaliella salina*. *J. Photochem. Photobiol. B: Biol.*, **87**: 137-143.

Choi, J., Joo, W. A., Park, S. J., Lee, S. H. and Kim, C. W. (2005) An efficient proteomics based strategy for the functional characterization of a novel halophilic enzyme from *Halobacterium salinarum*. *Proteomics.*, **5**: 907-917.

Christine, K. S., MacFarlane, A. W. t., Yang, K. and Stanley, R. J. (2002) Cyclobutylpyrimidine dimer base flipping by DNA photolyase. *J. Biol. Chem.*, **277**: 38339-38344.

Clark, A. E., Walsby A.E. (1978) The development and vertical distribution of populations of gas-vacuolate bacteria in a eutrophic, monomictic lake *Arch. Microbiol.*, **118**: 229-233.

Cline, S. D. and Hanawalt, P. C. (2003) Who's on first in the cellular response to DNA damage? *Nat. Rev. Mol. Cell. Biol.*, **4**: 361-372.

Cline, S. W., Lam, W. L., Charlebois, R. L., Schalkwyk, L. C. and Doolittle, W. F. (1989) Transformation methods for halophilic archaebacteria. *Can. J. Microbiol.*, **35**: 148-152.

Cline, S. W., Pfeifer, F. and Doolittle, W. F. (1995) Transformation of Halophilic Archaea. *Archaea: a laboratory manual.* F. T. Robb, K. R. Sowers, A. R. Place and H. J. Schreier, CSHL Press**:** 197-204.

Coesel, S., Mangogna, M., Ishikawa, T., Heijde, M., Rogato, A., Finazzi, G., Todo, T., Bowler, C. and Falciatore, A. (2009) Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. *EMBO Reports.*, **10**: 655-661.

Coker, J. A., DasSarma, P., Kumar, J., Muller, J. A. and DasSarma, S. (2007) Transcriptional profiling of the model Archaeon *Halobacterium* sp. NRC-1: responses to changes in salinity and temperature. *Saline Systems.*, **3**: 6.

Coker, J. A. and DasSarma, S. (2007) Genetic and transcriptomic analysis of transcription factor genes in the model halophilic Archaeon: coordinate action of TbpD and TfbA. *BMC. Genet.*, **8**: 61.

Connaris, H., Chaudhuri, J. B., Danson, M. J. and Hough, D. W. (1999) Expression, reactivation, and purification of enzymes from *Haloferax volcanii* in *Escherichia coli*. *Biotechnol. Bioeng.*, **64**: 38-45.

Cook, J. S. and Regan, J. D. (1969) Photoreactivation and photoreactivating enzyme activity in an order of mammals (Marsupialia). *Nature.*, **223**: 1066-1067.

Courcelle, J., Khodursky, A., Peter, B., Brown, P. O. and Hanawalt, P. C. (2001) Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics*. **158**: 41-64.

Crowley, D. J., Boubriak, I., Berquist, B. R., Clark, M., Richard, E., Sullivan, L., DasSarma, S. and McCready, S. (2006) The *uvrA*, *uvrB* and *uvrC* genes are required for repair of ultraviolet light induced DNA photoproducts in *Halobacterium* sp. NRC-1. *Saline Systems.*, **2**: 11.

Daiyasu, H., Ishikawa, T., Kuma, K., Iwai, S., Todo, T. and Toh, H. (2004) Identification of cryptochrome DASH from vertebrates. *Genes to Cells*. **9**: 479-495.

Dambeck, M. and Soppa, J. (2008) Characterization of a *Haloferax volcanii* member of the enolase superfamily: deletion mutant construction, expression analysis, and transcriptome comparison. *Arch. Microbiol.*, **190**: 341-353.

DasSarma, S., DasSarma P. (1997) Genetic analysis of the gas vesicle gene cluster in haloarchaea.*FEMS. Microbiol. Lett.*, **153**: 1-10.

DasSarma, S., et al. (2001) Genomic perspective on the photobiology of *Halobacterium* species NRC-1, a phototrophic, phototactic, and UV-tolerant haloarchaeon. *Photosynth. Res.*, **70**: 3-17.

DasSarma S. and Aurora, P. (2002) Halophiles. *Encyclopedia of Life Sciences*. London: Nature Publishing Group, Macmillan Publishers Ltd. **8:** 458-466.

De Bernardez Clark, E., Schwarz, E. and Rudolph, R. (1999) Inhibition of aggregation side reactions during *in vitro* protein folding. *Methods. Enzymol.*, **309**: 217-236.

De Bernardez-Clark, E. and Georgiou, G. (1991) *Inclusion bodies and recovery of proteins from the aggregated state.* . ACS Symposium Series 470, Washington DC, American Chemical Society.

De Rosa, M., Gambacorta, A. and Gliozzi, A. (1986) Structure, biosynthesis, and physicochemical properties of archaebacterial lipids. *Microbiol Rev*. **50**: 70-80.

Delsuc, F., Brinkmann, H. and Philippe, H. (2005) Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.*, **6**: 361-375.

DeRosa, M. C., Sancar, A. and Barton, J. K. (2005) Electrically monitoring DNA repair by photolyase. *Proc. Natl. Acad. Sci. U. S. A.*, **102**: 10788-10792.

Desmond, E., Brochier-Armanet, C. and Gribaldo, S. (2007) Phylogenomics of the archaeal flagellum: rare horizontal gene transfer in a unique motility structure. *BMC. Evol. Biol.*, **7**: 106.

DeVeaux, L. C., Muller, J. A., Smith, J., Petrisko, J., Wells, D. P. and DasSarma, S. (2007) Extremely radiation-resistant mutants of a halophilic archaeon with increased single-stranded DNA-binding protein (RPA) gene expression. *Radiat. Res.*, **168**: 507-514.

Diaz, S., Perez-Pomares, F., Pire, C., Ferrer, J. and Bonete, M. J. (2006) Gene cloning, heterologous overexpression and optimized refolding of the NAD-glutamate dehydrogenase from *Haloferax mediterranei*. *Extremophiles.*, **10**: 105-115.

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. and Mattick, J. S. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic. Acids. Res.*, **19**: 4008.

Dorazi, R., Gotz, D., Munro, S., Bernander, R. and White, M. F. (2007) Equal rates of repair of DNA photoproducts in transcribed and non-transcribed strands in *Sulfolobus solfataricus*. *Mol. Microbiol.*, **63**: 521-529.

Dundas, I. D. and Larsen, H. (1963) A Study on the killing by light of photosensitized cells of *Halobacterium salinarium*. *Arch. Mikrobiol.*, **46**: 19-28.

Dundas, I. E. (1977) Physiology of *Halobacteriaceae*. *Adv. Microb. Physiol.*, **15**: 85-120.

Dutta, K., Hejmadi, V. S. and Verma, N. C. (1993) An assay for DNA photolyases using non-radioactive DNA and restriction enzymes. *J. Photochem. Photobiol. B: Biol.*, **18**: 211-214.

Dyall-Smith, M. L. (2008) The halohandbook: protocols for halobacterial genetics [online] http://www.haloarchaea.com/resources/halohandbook/index.html.

Dym, O., Mevarech, M. and Sussman, J. L. (1995) Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. *Science.*, **267**: 1344-1346.

Edgell, D. R. and Doolittle, W. F. (1997) Archaea and the origin(s) of DNA replication proteins. *Cell.*, **89**: 995-998.

Eichler, J. (2001) Biotechnological uses of archaeal extremozymes. *Biotechnol. Adv.*, **19**: 261-278.

Eisenberg, H. and Wachtel, E. J. (1987) Structural studies of halophilic proteins, ribosomes, and organelles of bacteria adapted to extreme salt concentrations. *Annu Rev Biophys Biophys Chem*. **16**: 69-92.

Eker, A. P. (1980) Photoreactivating enzyme from *Streptomyces griseus*--II. Evidence for the presence of an intrinsic chromophore. *Photochem. Photobiol.*, **32**: 593-600.

Eker, A. P., Dekker, R. H. and Berends, W. (1981) Photoreactivating enzyme from *Streptomyces griseus*-IV. On the nature of the chromophoric cofactor in *Streptomyces griseus* photoreactivating enzyme. *Photochem. Photobiol.*, **33**: 65-72.

Eker, A. P. and Fichtinger-Schepman, A. M. (1975) Studies on a DNA photoreactivating enzyme from *Streptomyces griseus*. II. Purification of the enzyme. *Biochim. Biophys. Acta.*, **378**: 54-63.

Eker, A. P., Hessels, J. K. C. and Vandevelde, J. (1988) Photoreactivating Enzyme from the Green-Alga *Scenedesmus-Acutus* - Evidence for the Presence of 2 Different Flavin Chromophores. *Biochemistry.*, **27**: 1758-1765.

Eker, A. P. M., Formenoy, L. and Dewit, L. E. A. (1991) Photoreactivation in the extreme halophilic archaebacterium *Halobacterium cutirubrum*. *Photochem. Photobiol.*, **53**: 643-651.

Eker, A. P. M., Pol, A., Vandermeyden, P. and Vogels, G. D. (1980) Purification and properties of 8-hydroxy-5-deazaflavin derivatives from *Streptomyces griseus*. *FEMS. Microbiol. Lett.*, **8**: 161-165.

Eker, A. P. M., Yajima, H. and Yasui, A. (1994) DNA photolyase from the fungus *Neurospora crassa* - Purification, characterization and comparison with other photolyases. *Photochem. Photobiol.*, **60**: 125-133.

Elcock, A. H. and McCammon, J. A. (1998) Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.*, **280**: 731-748.

Elkins, J. G., et al. (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl. Acad. Sci. U. S. A.*, **105**: 8102-8107.

Falb, M., et al. (2006) Archaeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey. *J. Mol. Biol.*, **362**: 915-924.

Falb, M., Pfeiffer, F., Palm, P., Rodewald, K., Hickmann, V., Tittor, J. and Oesterhelt, D. (2005) Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis*. *Genome. Res.*, **15**: 1336-1343.

Ferrer, M., Lundsdorf, H., Chernikova, T. N., Yakimov, M., Timmis, K. N. and Golyshin, P. N. (2004) Functional consequences of single:double ring transitions in chaperonins: life in the cold. *Mol. Microbiol.*, **53**: 167-182.

Fine, A., Irihimovitch, V., Dahan, I., Konrad, Z. and Eichler, J. (2006) Cloning, expression, and purification of functional Sec11a and Sec11b, type I signal peptidases of the archaeon *Haloferax volcanii*. *J. Bacteriol.*, **188**: 1911-1919.

Fitt, P. S. and Sharma, N. (1987) The fate of thymine-containing dimers in ultraviolet-irradiated *Halobacterium cutirubrum*. *Biochim. Biophys. Acta.*, **910**: 103-110.

Fitt, P. S., Sharma, N. and Castellanos, G. (1983) A comparison of liquid-holding recovery and photoreactivation in halophilic and non-halophilic bacteria. *Biochim. Biophys. Acta.*, **739**: 73-78.

Forterre, P. (1999) Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. *Mol. Microbiol.*, **33**: 457-465.

Forterre, P. and Gadelle, D. (2009) Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic. Acids. Res.*, **37**: 679-692.

Forterre, P., Gribaldo, S. and Brochier-Armanet, C. (2009) Happy together: genomic insights into the unique *Nanoarchaeum/Ignicoccus* association. *J. Biol.*, **8**: 7.

Franzmann, D., Stackebrandt E., Sanderson K., Volkman J.K., Cameron D.E.,Stevenson P.L., McMeekin T.A. and H.R, B. (1988) *Halobacterium lacusprofundi* sp. nov., a halophilic bacterium isolated from Deep Lake, Antarctica. *Syst. Appl. Microbiol.* **11**: 20-27.

Friedberg, E., Walker, G. and Siede, W. (1995) *DNA repair and mutagenesis*, ASM Press.

Frolow, F., Harel, M., Sussman, J. L., Mevarech, M. and Shoham, M. (1996) Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. *Nat. Struct. Biol.*, **3**: 452-458.

Frols, S., Gordon, P. M., Panlilio, M. A., Duggin, I. G., Bell, S. D., Sensen, C. W. and Schleper, C. (2007) Response of the hyperthermophilic archaeon *Sulfolobus solfataricus* to UV damage. *J. Bacteriol.*, **189**: 8708-8718.

Frols, S., White, M. F. and Schleper, C. (2009) Reactions to UV damage in the model archaeon *Sulfolobus solfataricus*. *Biochem. Soc. Trans.*, **37**: 36-41.

Fujihashi, M., Numoto, N., Kobayashi, Y., Mizushima, A., Tsujimura, M., Nakamura, A., Kawarabayasi, Y. and Miki, K. (2007) Crystal structure of archaeal photolyase from *Sulfolobus tokodaii* with two FAD molecules: Implication of a novel light-harvesting cofactor. *J. Mol. Biol.*, **365**: 903-910.

Gao, B. and Gupta, R. S. (2007) Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis. *BMC. Genomics.*, **8**: 86.

Giampieri, S. and Storey, A. (2004) Repair of UV-induced thymine dimers is compromised in cells expressing the E6 protein from human papillomaviruses types 5 and 18. *Br. J. Cancer.*, **90**: 2203-2209.

Gibbons, N. E. (1974) Halobacteriaceae. . *Bergey's Manual of Determinative Bacteriology*. G. N. Buchanan RE. Baltimore, Williams & Wilkins.**:** 269-273.

Gogarten, J. P., et al. (1989) Evolution of the vacuolar H+-ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.*, **86**: 6661-6665.

Gomes, J. and Steiner, W. (2004) The biocatalytic potential of extremophiles and extremozymes. *Food. Technol. Biotech.*, **42**: 223-235.

Goo, Y. A., et al. (2004) Low-pass sequencing for microbial comparative genomics. *BMC. Genomics.*, **5**: 3.

Goodsell, D. S. (2005) Recognition in action: flipping pyrimidine dimers. *J. Mol. Recognit.*, **18**: 193-195.

Gotz, D., Paytubi, S., Munro, S., Lundgren, M., Bernander, R. and White, M. F. (2007) Responses of hyperthermophilic crenarchaea to UV irradiation. *Genome. Biol.*, **8**: R220.

Grant, W. D., Kamekura, M., McGenity, T. J. and Ventosa, A. (2001) Class III. Halobacteria class. nov. *Bergey's Manual of Systematic Bacteriology*. Boone D.R., Castenholz R.W. and G. G.M. New York, Springer. **1:** 294.

Grant, W. D. and Larsen, H. (1990) Extremely halophilic archaeobacteria, order halobacteriales ord.nov. . *Bergey's manual of systermic bacteriology*. B. M. Stanley JT, Pfennig N, Holt JG. **3:** 2216-2233.

Grey, V. L. and Fitt, P. S. (1976) Evidence for the lack of deoxyribonucleic acid dark-repair in *Halobacterium cutirubrum*. *Biochem. J.*, **156**: 569-575.

Gribaldo, S. and Brochier-Armanet, C. (2006) The origin and evolution of Archaea: a state of the art. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **361**: 1007-1022.

Gribaldo, S. and Cammarano, P. (1998) The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *J. Mol. Evol.*, **47**: 508-516.

Grogan, D. W. (2000) The question of DNA repair in hyperthermophilic archaea. *Trends. Microbiol.*, **8**: 180-185.

Gu, Z., Weidenhaupt, M., Ivanova, N., Pavlov, M., Xu, B., Su, Z. G. and Janson, J. C. (2002) Chromatographic methods for the isolation of, and refolding of proteins from, *Escherichia coli* inclusion bodies. *Protein. Expr. Purif.*, **25**: 174-179.

Guo, H., Yang, H., Mockler, T. C. and Lin, C. (1998) Regulation of flowering time by *Arabidopsis* photoreceptors. *Science.*, **279**: 1360-1363.

Ha, D. J., Joo, W. A., Han, G. Y. and Kim, C. W. (2007) Proteome analysis of *Halobacterium salinarum* and characterization of proteins related to the degradation of isopropyl alcohol. *Biochim. Biophys. Acta.*, **1774**: 44-50.

Hamm-Alvarez, S., Sancar, A. and Rajagopalan, K. V. (1989) Role of enzyme-bound 5,10-methenyltetrahydropteroylpolyglutamate in catalysis by *Escherichia coli* DNA photolyase. *J. Biol. Chem.*, **264**: 9649-9656.

Hamm-Alvarez, S., Sancar, A. and Rajagopalan, K. V. (1990) The folate cofactor of *Escherichia coli* DNA photolyase acts catalytically. *J. Biol. Chem.*, **265**: 18656-18662.

Hamm-Alvarez, S. F., Sancar, A. and Rajagopalan, K. V. (1990) The presence and distribution of reduced folates in *Escherichia coli* dihydrofolate reductase mutants. *J. Biol. Chem.*, **265**: 9850-9856.

Harm, H. and Rupert, C. S. (1970a) Analysis of photoenzymatic repair of UV lesions in DNA by single light flashes. VI. Rate constants for enzyme-substrate binding in vitro between yeast photoreactivating enzyme and ultraviolet lesions in *Haemophilus* transforming DNA. *Mutat. Res.*, **10**: 291-306.

Harm, H. and Rupert, C. S. (1970b) Analysis of photoenzymatic repair of UV lesions in DNA by single light flashes. VII. Photolysis of enzyme-substrate complexes *in vitro*. *Mutat. Res.*, **10**: 307-318.

Harm, W. (1970) Analysis of photoenzymatic repair of UV lesions in DNA by single light flashes. V. Determination of the reaction-rate constants in *E. coli* cells. *Mutat. Res.*, **10**: 277-290.

Harm, W., Harm, H. and Rupert, C. S. (1968) Analysis of photoenzymatic repair of UV lesions in DNA by single light flashes. II. *In vivo* studies with *Escherichia coli* cells and bacteriophage. *Mutat. Res.*, **6**: 371-385.

Hartmann, R., Sickinger, H. D. and Oesterhelt, D. (1980) Anaerobic growth of halobacteria. *Proc. Natl. Acad. Sci. U. S. A.*, **77**: 3821-385.

Hescox, M. A. and Carlberg, D. M. (1972) Photoreactivation in *Halobacterium cutirubrum*. *Can. J. Microbiol.*, **18**: 981-985.

Hirouchi, T., et al. (2003) A gene for a Class II DNA photolyase from *Oryza sativa*: cloning of the cDNA by dilution-amplification. *Mol. Genet. Genomics.* . **269**: 508-516.

Hitomi, K., et al. (2009) Functional motifs in the (6-4) photolyase crystal structure make a comparative framework for DNA repair photolyases and clock cryptochromes. *Proc. Natl. Acad. Sci. U. S. A.*, **106**: 6962-6967.

Hitomi, K., Okamoto, K., Daiyasu, H., Miyashita, H., Iwai, S., Toh, H., Ishiura, M. and Todo, T. (2000) Bacterial cryptochrome and photolyase: characterization of two photolyase-like genes of *Synechocystis* sp PCC6803. *Nucleic. Acids. Res.*, **28**: 2353-2362.

Holmes, M., Pfeifer, F. and Dyall-Smith, M. (1994) Improved shuttle vectors for *Haloferax volcanii* including a dual-resistance plasmid. *Gene.*, **146**: 117-121.

Holmes, M. L. and Dyall-Smith, M. L. (1990) A plasmid vector with a selectable marker for halophilic archaebacteria. *J. Bacteriol.*, **172**: 756-761.

Holmes, M. L. and Dyall-Smith, M. L. (2000) Sequence and expression of a halobacterial *beta-galactosidase* gene. *Mol. Microbiol.*, **36**: 114-122.

Holmes, M. L., Nuttall, S. D. and Dyall-Smith, M. L. (1991) Construction and use of halobacterial shuttle vectors and further studies on *Haloferax* DNA gyrase. *J. Bacteriol.*, **173**: 3807-3813.

Hough, D. W. and Danson, M. J. (1999) Extremozymes. *Curr. Opin. Chem. Biol.*, **3**: 39-46.

Houwink, A. L. (1956) Flagella, gas vacuoles and cell-wall structure in *Halobacterium halobium*; an electron microscope study. *J. Gen. Microbiol.*, **15**: 146-150.

Howard-Flanders, P., Boyce, R. P. and Theriot, L. (1966) Three loci in *Escherichia coli* K-12 that control the excision of pyrimidine dimers and certain other mutagen products from DNA. *Genetics.*, **53**: 1119-1136.

Hsu, D. S., Zhao, X., Zhao, S., Kazantsev, A., Wang, R. P., Todo, T., Wei, Y. F. and Sancar, A. (1996) Putative human blue-light photoreceptors hCRY1 and hCRY2 are flavoproteins. *Biochemistry.*, **35**: 13871-13877.

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C. and Stetter, K. O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature.*, **417**: 63-67.

Huber, H., Hohn, M. J., Stetter, K. O. and Rachel, R. (2003) The phylum *Nanoarchaeota*: present knowledge and future perspectives of a unique form of life. *Res. Microbiol.*, **154**: 165-171.

Huber, R., Huber, H. and Stetter, K. O. (2000) Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. *FEMS. Microbiol. Rev.*, **24**: 615-623.

Husain, I., Sancar, G. B., Holbrook, S. R. and Sancar, A. (1987) Mechanism of damage recognition by *Escherichia coli* DNA photolyase. *J. Biol. Chem.*, **262**: 13188-13197.

Ishibashi, M., Tokunaga, H., Hiratsuka, K., Yonezawa, Y., Tsurumaru, H., Arakawa, T. and Tokunaga, M. (2001) NaCl-activated nucleoside diphosphate kinase from extremely halophilic archaeon, *Halobacterium salinarum*, maintains native conformation without salt. *FEBS. Lett.*, **493**: 134-138.

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. and Miyata, T. (1989) Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. U. S. A.*, **86**: 9355-9359.

Iwasa, T., Tokutomi, S. and Tokunaga, F. (1988) Photoreactivation of *Halobacterium halobium* - Action ppectrum and role of pigmentation. *Photochem. Photobiol.*, **47**: 267-270.

Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K. O., Rachel, R. and Huber, H. (2008) *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: New insights into a unique, intimate association of two archaea. *J. Bacteriol.*, **190**: 1743-1750.

Jain, R., Rivera, M. C. and Lake, J. A. (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proc. Natl. Acad. Sci. U. S. A.*, **96**: 3801-3806.

Jans, J., et al. (2005) Powerful skin cancer protection by a CPD-photolyase transgene. *Curr. Biol.*, **15**: 105-115.

Johnson, J. L., Hammalvarez, S., Payne, G., Sancar, G. B., Rajagopalan, K. V. and Sancar, A. (1988) Identification of the 2nd Chromophore of *Escherichia coli* and Yeast DNA Photolyases as 5,10-Methenyltetrahydrofolate. *Proc. Natl. Acad. Sci. U. S. A.*, **85**: 2046-2050.

Jolley, K. A., Rapaport, E., Hough, D. W., Danson, M. J., Woods, W. G. and Dyall-Smith, M. L. (1996) Dihydrolipoamide dehydrogenase from the halophilic archaeon *Haloferax volcanii*: homologous overexpression of the cloned gene. *J. Bacteriol.*, **178**: 3044-3048.

Jorns, M. S., Baldwin, E. T., Sancar, G. B. and Sancar, A. (1987) Action mechanism of *Escherichia coli* DNA photolyase. 2. Role of the chromophores in catalysis. *J. Biol. Chem.*, **262**: 486-491.

Jorns, M. S., Sancar, G. B. and Sancar, A. (1984) Identification of a neutral flavin radical and characterization of a 2nd chromophore in *Escherichia coli* DNA photolyase. *Biochemistry.*, **23**: 2673-2679.

Jorns, M. S., Wang, B. Y., Jordan, S. P. and Chanderkar, L. P. (1990) Chromophore function and interaction in *Escherichia coli* DNA photolyase: reconstitution of the apoenzyme with pterin and/or flavin derivatives. *Biochemistry.*, **29**: 552-561.

Kamekura, M., Dyall-Smith, M. L., Upasani, V., Ventosa, A. and Kates, M. (1997) Diversity of alkaliphilic halobacteria: proposals for transfer of *Natronobacterium vacuolatum*, *Natronobacterium magadii,* and *Natronobacterium pharaonis* to *Halorubrum*, *Natrialba*, and *Natronomonas* gen. nov., respectively, as *Halorubrum vacuolatum* comb. nov., *Natrialba magadii* comb. nov., and *Natronomonas pharaonis* comb. nov., respectively. *Int. J. Syst. Bacteriol.*, **47**: 853-857.

Kamekura, M. and Seno, Y. (1990) A halophilic extracellular protease from a halophilic archaebacterium strain 172 P1. *Biochem. Cell. Biol.*, **68**: 352-359.

Kane, J. F. and Hartley, D. L. (1988) Formation of recombinant protein inclusion bodies in *Escherichia coli*. *Trends. Biotechnol.* . **6**: 95-101.

Kao, Y. T., Saxena, C., Wang, L., Sancar, A. and Zhong, D. (2005) Direct observation of thymine dimer repair in DNA by photolyase. *Proc. Natl. Acad. Sci. U. S. A.*, **102**: 16128-16132.

Kates, M. (1993) Biology of halophilic bacteria, Part II. Membrane lipids of extreme halophiles: biosynthesis, function and evolutionary significance. *Experientia.*, **49**: 1027-1036.

Kato, R., Hasegawa, K., Hidaka, Y., Kuramitsu, S. and Hoshino, T. (1997) Characterization of a thermostable DNA photolyase from an extremely thermophilic bacterium, *Thermus thermophilus* HB27. *J. Bacteriol.*, **179**: 6499-6503.

Kato, T., Jr., Todo, T., Ayaki, H., Ishizaki, K., Morita, T., Mitra, S. and Ikenaga, M. (1994) Cloning of a marsupial DNA photolyase gene and the lack of related nucleotide sequences in placental mammals. *Nucleic. Acids. Res.*, **22**: 4119-4124.

Kaur, A., Pan, M., Meislin, M., Facciotti, M. T., El-Gewely, R. and Baliga, N. S. (2006) A systems view of haloarchaeal strategies to withstand stress from transition metals. *Genome. Res.*, **16**: 841-854.

Keeling, P. J., Charlebois, R. L. and Doolittle, W. F. (1994) Archaebacterial genomes: eubacterial form and eukaryotic content. *Curr. Opin. Genet. Dev.*, **4**: 816-822.

Kelman, Z. and White, M. F. (2005) Archaeal DNA replication and repair. *Curr. Opin. Microbiol.*, **8**: 669-676.

Kelner, A. (1949) Effect of visible light on the recovery of *Streptomyces griseus conidia* from ultra-violet irradiation injury. *Proc. Natl. Acad. Sci. U. S. A.*, **35**: 73-79.

Kelner, A. (1949) Photoreactivation of ultraviolet-irradiated *Escherichia coli*, with special reference to the dose-reduction principle and to ultraviolet-induced mutation. *J. Bacteriol.*, **58**: 511-522.

Kelner, A. (1951) Action spectra for photoreactivation of ultraviolet-irradiated *Escherichia coli* and *Streptomyces griseus*. *J. Gen. Physiol.*, **34**: 835-852.

Kim, S. and Lee, S. B. (2006) Rare codon clusters at 5 '-end influence heterologous expression of archaeal gene in *Escherichia coli*. *Protein. Expr. Purif.*, **50**: 49-57.

Kim, S. T., Malhotra, K., Smith, C. A., Taylor, J. S. and Sancar, A. (1994) Characterization of (6-4)-Photoproduct DNA photolyase. *J. Biol. Chem.*, **269**: 8535-8540.

Kim, S. T., Malhotra, K., Taylor, J. S. and Sancar, A. (1996) Purification and partial characterization of (6-4) photoproduct DNA photolyase from *Xenopus laevis*. *Photochem. Photobiol.*, **63**: 292-295.

Kirkland, P. A., Gil, M. A., Karadzic, I. M. and Maupin-Furlow, J. A. (2008) Genetic and proteomic analyses of a proteasome-activating nucleotidase A mutant of the haloarchaeon *Haloferax volcanii*. *J. Bacteriol.*, **190**: 193-205.

Kirkland, P. A., Reuter, C. J. and Maupin-Furlow, J. A. (2007) Effect of proteasome inhibitor clasto-lactacystin-beta-lactone on the proteome of the haloarchaeon *Haloferax volcanii*. *Microbiology.*, **153**: 227122-80.

Klein, C., et al. (2007) The low molecular weight proteome of *Halobacterium salinarum*. *J. Proteome. Res.*, **6**: 1510-1518.

Klein, C., Garcia-Rizo, C., Bisle, B., Scheffer, B., Zischka, H., Pfeiffer, F., Siedler, F. and Oesterhelt, D. (2005) The membrane proteome of *Halobacterium salinarum*. *Proteomics.*, **5**: 180-197.

Kleine, T., Lockhart, P. and Batschauer, A. (2003) An *Arabidopsis* protein closely related to *Synechocystis* cryptochrome is targeted to organelles. *Plant. J.*, **35**: 93-103.

Kobayashi, K., Kanno, S., Smit, R., van der Horst, G. T. J., Takao, M. and Yasui, A. (1998) Characterization of photolyase/blue-light receptor homologs in mouse and human cells. *Nucleic. Acids. Res.*, **26**: 5086-5092.

Komori, H., Masui, R., Kuramitsu, S., Yokoyama, S., Shibata, T., Inoue, Y. and Miki, K. (2001) Crystal structure of thermostable DNA photolyase: pyrimidine-dimer recognition mechanism. *Proc. Natl. Acad. Sci. U. S. A.*, **98**: 13560-13565.

Koonin, E. V., Mushegian, A. R., Galperin, M. Y. and Walker, D. R. (1997) Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol. Microbiol.*, **25**: 619-637.

Kort, R., Komori, H., Adachi, S., Miki, K. and Eker, A. (2004) DNA apophotolyase from *Anacystis nidulans*: 1.8 angstrom structure, 8-HDF reconstitution and X-ray-induced FAD reduction. *Acta. Crystallogr. D. Biol. Crystallogr.* . **60**: 1205-1213.

Krishnan, D. and Painter, R. B. (1973) Photoreactivation and repair replication in rat kangaroo cells. *Mutat. Res.*, **17**: 213-222.

Kundu, L. M., Burgdorf, L. T., Kleiner, O., Batschauer, A. and Carell, T. (2002) Cleavable substrate containing molecular beacons for the quantification of DNA-photolyase activity. *Chembiochem.*, **3**: 1053-1060.

Labedan, B., et al. (1999) The evolutionary history of carbamoyltransferases: A complex set of paralogous genes was already present in the last universal common ancestor. *J. Mol. Evol.*, **49**: 461-473.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.*, **227**: 680-685.

Lam, W. L. and Doolittle, W. F. (1989) Shuttle vectors for the archaebacterium *Halobacterium volcanii*. *Proc. Natl. Acad. Sci. U. S. A.*, **86**: 5478-5482.

Lanyi, J. K. (1974) Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol. Rev.*, **38**: 272-290.

Large, A., Stamme, C., Lange, C., Duan, Z. H., Allers, T., Soppa, J. and Lund, P. A. (2007) Characterization of a tightly controlled promoter of the halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential cct1 gene. *Mol. Microbiol.*, **66**: 1092-1106.

LaVallie, E., DiBlasio, E., Kovacic, S., Grant, K., Schendel, P. and McCoy, J. (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology (N Y)*. **11**: 187-193.

Lawson, F. S., Charlebois, R. L. and Dillon, J. A. (1996) Phylogenetic analysis of *carbamoylphosphate synthetase* genes: complex evolutionary history includes an internal duplication within a gene which can root the tree of life. *Mol. Biol. Evol.*, **13**: 970-977.

Lechner, J. and Sumper, M. (1987) The primary structure of a procaryotic glycoprotein. Cloning and sequencing of the cell surface glycoprotein gene of halobacteria. *J. Biol. Chem.*, **262**: 9724-9729.

Lechner, J. and Wieland, F. (1989) Structure and biosynthesis of prokaryotic glycoproteins. *Annu. Rev. Biochem.*, **58**: 173-194.

Leipe, D. D., Aravind, L. and Koonin, E. V. (1999) Did DNA replication evolve twice independently? *Nucleic. Acids. Res.*, **27**: 3389-3401.

Levin, I., Giladi, M., Altman-Price, N., Ortenberg, R. and Mevarech, M. (2004) An alternative pathway for reduced folate biosynthesis in bacteria and halophilic archaea. *Mol. Microbiol.*, **54**: 1307-1318.

Li, Y. F. and Sancar, A. (1991) Cloning, sequencing, expression and characterization of DNA photolyase from *Salmonella typhimurium*. *Nucleic. Acids. Res.*, **19**: 4885-4890.

Lucas-Lledo, J. I. and Lynch, M. (2009) Evolution of mutation rates: phylogenomic analysis of the photolyase/cryptochrome family. *Mol. Biol. Evol.*, **26**: 1143-1153.

Macelroy, R. D. (1974) Some comments on evolution of extremophiles. *Biosystems.*, **6**: 74-75.

Madern, D., Pfister, C. and Zaccai, G. (1995) Mutation at a single acidic amino acid enhances the halophilic behaviour of malate dehydrogenase from *Haloarcula marismortui* in physiological salts. *Eur. J. Biochem.*, **230**: 1088-1095.

Madigan, M. T. and Marrs, B. L. (1997) Extremophiles. *Sci. Am.*, **276**: 82-87.

Malhotra, K., Baer, M., Li, Y. F., Sancar, G. B. and Sancar, A. (1992) Identification of chromophore binding domains of yeast DNA photolyase. *J. Biol. Chem.*, **267**: 2909-2914.

Malhotra, K., Kim, S. T., Batschauer, A., Dawut, L. and Sancar, A. (1995) Putative blue-light photoreceptors from *Arabidopsis thaliana* and *Sinapis alba* with a high degree of sequence homology to DNA photolyase contain the two photolyase cofactors but lack DNA repair activity. *Biochemistry.*, **34**: 6892-6899.

Malhotra, K., Kim, S. T., Walsh, C. and Sancar, A. (1992) Roles of FAD and 8-hydroxy-5-deazaflavin chromophores in photoreactivation by *Anacystis nidulans* DNA photolyase. *J. Biol. Chem.*, **267**: 15406-15411.

Mancinelli, R. L. and Hochstein, L. I. (1986) The occurrence of denitrification in extremely halophilic bacteria. *FEMS. Microbiol. Lett.*, **35**: 55-58.

Margesin, R. and Schinner, F. (2001) Potential of halotolerant and halophilic microorganisms for biotechnology. *Extremophiles.*, **5**: 73-83.

Marinus, M. G. (1987) DNA methylation in *Escherichia coli*. *Annu. Rev. Genet.*, **21**: 113-131.

Marinus, M. G. and Morris, N. R. (1973) Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12. *J. Bacteriol.*, **114**: 1143-1150.

Masson, F., Laino, T., Rothlisberger, U. and Hutter, J. (2009) A QM/MM investigation of thymine dimer radical anion splitting catalyzed by DNA photolyase. *Chemphyschem.*, **10**: 400-410.

Maul, M. J., Barends, T. R., Glas, A. F., Cryle, M. J., Domratcheva, T., Schneider, S., Schlichting, I. and Carell, T. (2008) Crystal structure and mechanism of a DNA (6-4) photolyase. *Angew. Chem. Int. Ed. Engl.*, **47**: 10076-10080.

May, M. S. and Hattman, S. (1975) Analysis of bacteriophage deoxyribonucleic-acid sequences methylated by host-controlled and R-factor-controlled enzymes. *J. Bacteriol.*, **123**: 768-770.

McCready, S. (1996) The repair of ultraviolet light-induced DNA damage in the halophilic archaebacteria, *Halobacterium cutirubrum, Halobacterium halobium* and *Haloferax volcanii*. *Mutat. Res.*, **364**: 25-32.

McCready, S. (1999) A dot blot immunoassay for UV photoproducts. *Methods. Mol. Biol.*, **113**: 147-156.

McCready, S. and Marcello, L. (2003) Repair of UV damage in *Halobacterium salinarum*. *Biochem. Soc. Trans.*, **31**: 694-698.

McCready, S., Muller, J. A., Boubriak, I., Berquist, B. R., Ng, W. L. and DasSarma, S. (2005) UV irradiation induces homologous recombination genes in the model archaeon, *Halobacterium* sp. NRC-1. *Saline Systems.*, **1**: 3.

McCullough, A. K., Dodson, M. L., Scharer, O. D. and Lloyd, R. S. (1997) The role of base flipping in damage recognition and catalysis by T4 endonuclease V. *J. Biol. Chem.*, **272**: 27210-27217.

McGenity, T. J., Gemmell, R. T. and Grant, W. D. (1998) Proposal of a new halobacterial genus *Natrinema* gen. nov., with two species *Natrinema pellirubrum* nom. nov. and *Natrinema pallidum* nom. nov. *Int. J. Syst. Bacteriol.*, **48 Pt 4**: 1187-1196.

Mees, A., Klar, T., Gnau, P., Hennecke, U., Eker, A. P., Carell, T. and Essen, L. O. (2004) Crystal structure of a photolyase bound to a CPD-like DNA lesion after *in situ* repair. *Science.*, **306**: 1789-1793.

Mesbah, N. M., Abou-El-Ela, S. H. and Wiegel, J. (2007) Novel and unexpected prokaryotic diversity in water and sediments of the alkaline, hypersaline lakes of the Wadi An Natrun, Egypt. *Microb. Ecol.*, **54**: 598-617.

Mescher, M. F. and Strominger, J. L. (1976) Structural (shape-maintaining) role of the cell surface glycoprotein of *Halobacterium salinarium*. *Proc. Natl. Acad. Sci. U. S. A.*, **73**: 2687-2691.

Messner, P. (1997) Bacterial glycoproteins. *Glycoconj. J.*, **14**: 3-11.

Mevarech, M. and Werczberger, R. (1985) Genetic transfer in *Halobacterium volcanii*. *J. Bacteriol.*, **162**: 461-462.

Montalvo-Rodriguez, R., Vreeland, R. H., Oren, A., Kessel, M., Betancourt, C. and Lopez-Garriga, J. (1998) *Halogeometricum borinquense* gen. nov., sp. nov., a novel halophilic archaeon from Puerto Rico. *Int. J. Syst. Bacteriol.*, **48 Pt 4**: 1305-1312.

Mullakhanbhai, M. F. and Larsen, H. (1975) *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch. Microbiol.*, **104**: 207-214.

Muller, J. A. and DasSarma, S. (2005) Genomic analysis of anaerobic respiration in the archaeon *Halobacterium* sp. strain NRC-1: dimethyl sulfoxide and trimethylamine N-oxide as terminal electron acceptors. *J. Bacteriol.*, **187**: 1659-1667.

Nakajima, S., et al. (1998) Cloning and characterization of a gene (UVR3) required for photorepair of 6-4 photoproducts in *Arabidopsis thaliana*. *Nucleic. Acids. Res.*, **26**: 638-644.

Nakayama, T., Todo, T., Notsu, S., Nakazono, M. and Zaitsu, K. (2004) Assay method for *Escherichia coli* photolyase activity using single-strand cis-syn cyclobutane pyrimidine dimer DNA as substrate. *Anal. Biochem.*, **329**: 263-268.

Ng, W. V., et al. (2000) Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. U. S. A.*, **97**: 12176-12181.

Norais, C., Hawkins, M., Hartman, A. L., Eisen, J. A., Myllykallio, H. and Allers, T. (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS. Genet.*, **3**: e77.

Novy, R., Drott, D., Yaeger, K. and Mierendorf, R. (2001) Overcoming the codon bia of *E. coli* for enhanced protein expression. *inNovations.*, **12**: 1-3.

Nygren, P., Ståhl, S. and Uhlén, M. (1994) Engineering proteins to facilitate bioprocessing. *Trends. Biotechnol.*, **12**: 184-188.

O'Connor, K. A., McBride, M. J., West, M., Yu, H., Trinh, L., Yuan, K., Lee, T. and Zusman, D. R. (1996) Photolyase of *Myxococcus xanthus*, a Gram-negative eubacterium, is more similar to photolyases found in archaea and "higher" eukaryotes than to photolyases of other eubacteria. *J. Biol. Chem.*, **271**: 6252-6259.

Oesterhelt, D. and Stoeckenius, W. (1971) Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nat. New. Biol.*, **233**: 149-152.

Oesterhelt, D. and Stoeckenius, W. (1973) Functions of a new photoreceptor membrane. *Proc. Natl. Acad. Sci. U. S. A.*, **70**: 2853-2857.

Offner, S. and Pfeifer, F. (1995) Complementation studies with the gas vesicle-encoding p-vac region of *Halobacterium salinarium* PHH1 reveal a regulatory role for the *p-gvpDE* genes. *Mol. Microbiol.*, **16**: 9-19.

Ogrunc, M., Becker, D. F., Ragsdale, S. W. and Sancar, A. (1998) Nucleotide excision repair in the third kingdom. *J. Bacteriol.*, **180**: 5796-5798.

Oren, A. (1999) Bioenergetic aspects of halophilism. *Microbiol. Mol. Biol. Rev.*, **63**: 334-348.

Oren, A. (2001) The Order Halobacteriales. *The Prokaryotes. A Handbook on the Biology of Bacteria:ecophysiology, isolation, identification, applications*. S. F. M. Dworkin, E. Rosenberg, K.-H. Schleifer, and E. Stackenbrandt. New York, Springer-Verlag**:** 2, 16-19.

Oren, A. (2002) Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications. *J. Ind. Microbiol. Biotechnol.*, **28**: 56-63.

Oren, A. (2006) Life at High Salt Concentrations. *The Prokaryotes: A Handbook on the Biology of Bacteria*. S. F. MM Dworkin, E Rosenberg, K-H Schleifer and E Stackebrandt. **2:** 263-282.

Oren, A., Arahal, D. R. and Ventosa, A. (2009) Emended descriptions of genera of the family *Halobacteriaceae. Int. J. Syst. Evol. Microbiol.*, **59**: 637-642.

Oren, A. and Truper, H. G. (1990) Anaerobic growth of halophilic archaeobacteria by reduction of dimethysulfoxide and trimethylamine N-oxide. *FEMS. Microbiol. Lett.*, **70**: 33-36.

Oren, A. and Ventosa, A. (2000) International Committee on Systematic Bacteriology Subcommittee on the Taxonomy of *Halobacteriaceae*. Minutes of the Meeting. *International Journal of Systematic and Evolutionary Microbiology*. Sydney, Australia. **50:** 1405-1407.

Ozturk, N., Kao, Y. T., Selby, C. P., Kavakli, I. H., Partch, C. L., Zhong, D. and Sancar, A. (2008) Purification and characterization of a type III photolyase from *Caulobacter crescentus*. *Biochemistry.*, **47**: 10255-10261.

Ozturk, N., Song, S. H., Ozturk, S., Selby, C. P., Morrison, L., Partch, C., Zhong, D. and Sancar, A. (2007) Structure and function of animal cryptochromes. *Cold. Spring. Harb. Symp. Quant. Biol.*, **72**: 119-131

Park, H. W., Kim, S. T., Sancar, A. and Deisenhofer, J. (1995) Crystal structure of DNA photolyase from *Escherichia coli*. *Science.*, **268**: 1866-1872.

Park, H. W., Sancar, A. and Deisenhofer, J. (1993) Crystallization and preliminary crystallographic analysis of *Escherichia coli* DNA photolyase. *J. Mol. Biol.*, **231**: 1122-1125.

Patenge, N., Haase, A., Bolhuis, H. and Oesterhelt, D. (2000) The gene for a halophilic beta-galactosidase (bgaH) of *Haloferax alicantei* as a reporter gene for promoter analyses in *Halobacterium salinarum*. *Mol. Microbiol.*, **36**: 105-113.

Payne, G. and Sancar, A. (1990) Absolute Action Spectrum of E-Fadh2 and E-Fadh2-Mthf Forms of *Escherichia coli* DNA Photolyase. *Biochemistry.*, **29**: 7715-7727.

Payne, G., Wills, M., Walsh, C. and Sancar, A. (1990) Reconstitution of *Escherichia coli* photolyase with flavins and flavin analogues. *Biochemistry.*, **29**: 5706-5711.

Petersen, J. L., Lang, D. W. and Small, G. D. (1999) Cloning and characterization of a class II DNA photolyase from *Chlamydomonas*. *Plant. Mol. Biol.*, **40**: 1063-1071.

Pettijohn, D. and Hanawalt, P. (1964) Evidence for repair-replication of ultraviolet damaged DNA in bacteria. *J. Mol. Biol.*, **9**: 395-410.

Pfeifer, F., Griffig, J. and Oesterhelt, D. (1993) The *fdx* gene encoding the [2Fe--2S] ferredoxin of *Halobacterium salinarium* (*H. halobium*). *Mol. Gen. Genet.*, **239**: 66-71.

Pfeiffer, F., et al. (2008) Evolution in the laboratory: the genome of *Halobacterium salinarum* strain R1 compared to that of strain NRC-1. *Genomics.*, **91**: 335-346.

Pieper, U., Kapadia, G., Mevarech, M. and Herzberg, O. (1998) Structural features of halophilicity derived from the crystal structure of dihydrofolate reductase from the Dead Sea halophilic archaeon, *Haloferax volcanii*. *Structure.*, **6**: 75-88.

Podar, M., et al. (2008) A genomic analysis of the archaeal system *Ignicoccus hospitalis-Nanoarchaeum equitans*. *Genome. Biol.*, **9**: R158.1-R158.18.

Prakash, S. and Prakash, L. (2000) Nucleotide excision repair in yeast. *Mutat. Res. Fundam. Mol. Mech. Mugag.*, **451**: 13-24.

Prangishvilli, D., Zillig, W., Gierl, A., Biesert, L. and Holz, I. (1982) DNA-dependent RNA polymerase of thermoacidophilic archaebacteria. *Eur. J. Biochem.*, **122**: 471-477.

Pryor, K. D. and Leiting, B. (1997) High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system. *Protein. Expr. Purif.*, **10**: 309-319.

Quillardet, P., Rouffaud, M. A. and Bouige, P. (2003) DNA array analysis of gene expression in response to UV irradiation in *Escherichia coli*. *Res. Microbiol.*, **154**: 559-572.

Ramage, P., Hemmig R., Mathis, B., Cowan-Jacob, S. W., Rondeau, J. M., Kallen, J., Blommer, M. J. J., Zurini, M. and Rudisser, S. (2002) Snags with tags: Some observations made with $(His)_6$-tagged proteins. *Innovations Forum.*1-4.

Rao, J. K. and Argos, P. (1981) Structural stability of halophilic proteins. *Biochemistry.*, **20**: 6536-6543.

Robinson, J. L., Pyzyna, B., Atrasz, R. G., Henderson, C. A., Morrill, K. L., Burd, A. M., Desoucy, E., Fogleman, R. E., 3rd., Naylor, J. B., Steele, S. M., Elliott, D. R., Leyva, K. J., and Shand, R. F. (2005) Growth kinetics of extremely halophilic archaea (family halobacteriaceae) as revealed by arrhenius plots. *J. Bacteriol.*, **187**: 923-929.

Robinson, N. P. and Bell, S. D. (2007) Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc. Natl. Acad. Sci. U. S. A.*, **104**: 5806-5811.

Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R. and Bell, S. D. (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell.*, **116**: 25-38.

Rodriguez-Valera, F. (1988) Characteristics and microbial ecology of hypersaline environments. *Halophilic bacteria*. F. Rodriguez-Valera. Florida, CRC Press Inc**:** 3-30.

Roh, S. W. and Bae, J. W. (2009) *Halorubrum cibi* sp. nov., an extremely halophilic archaeon from salt-fermented seafood. *J. Microbiol.*, **47**: 162-166.

Rokas, A., Williams, B. L., King, N. and Carroll, S. B. (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature.*, **425**: 798-804.

Romano, V., Napoli, A., Valenti, V. S. A., Rossi, M. and Ciaramella, M. (2007) Lack of strand-specific repair of UV-induced DNA lesions in three genes of the archaeon *Sulfolobus solfataricus*. *J. Mol. Biol.*, **365**: 921-929.

Rothschild, L. J. and Mancinelli, R. L. (2001) Life in extreme environments. *Nature.*, **409**: 1092-1101.

Ruepp, A. and Soppa, J. (1996) Fermentative arginine degradation in *Halobacterium salinarium* (formerly *Halobacterium halobium*): genes, gene products, and transcripts of the *arcRACB* gene cluster. *J. Bacteriol.*, **178**: 4942-4947.

Rupert, C. S., Goodgal, S. H. and Herriott, R. M. (1958) Photoreactivation *in vitro* of ultraviolet-inactivated *Hemophilus influenzae* transforming factor. *J. Gen. Physiol.*, **41**: 451-471.

Salerno, V., Napoli, A., White, M. F., Rossi, M. and Ciaramella, M. (2003) Transcriptional response to DNA damage in the archaeon *Sulfolobus solfataricus*. *Nucleic. Acids. Res.*, **31**: 6127-6138.

Sambrook, J. and Russell, D. (2001) *Molecular cloning: a laboratory manual*
New York, CSHL Press.

Sancar, A. (1994) Structure and function of DNA photolyase. *Biochemistry.*, **33**: 2-9.

Sancar, A. (2003) Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem. Rev.*, **103**: 2203-2237.

Sancar, A., Clarke, N. D., Griswold, J., Kennedy, W. J. and Rupp, W. D. (1981b) Identification of the *Uvrb*-Gene Product. *J. Mol. Biol.*, **148**: 63-76.

Sancar, A., Kacinski, B. M., Mott, D. L. and Rupp, W. D. (1981a) Identification of the U*uvrc*-gene product. *Proc. Nat. Acad. Sci. U.S.A.*, **78**: 5450-5454.

Sancar, A. and Rupp, W. D. (1983) A novel repair enzyme - Uvrabc excision nuclease of *Escherichia coli* cuts a DNA strand on both sides of the damaged region. *Cell.*, **33**: 249-260.

Sancar, A. and Sancar, G. B. (1984) *Escherichia coli* DNA photolyase is a flavoprotein. *J. Mol. Biol.*, **172**: 223-227.

Sancar, A., Smith, F. W. and Sancar, G. B. (1984a) Purification of *Escherichia coli* DNA photolyase. *J. Biol. Chem.*, **259**: 6028-6032.

Sancar, A., Wharton, R. P., Seltzer, S., Kacinski, B. M., Clarke, N. D. and Rupp, W. D. (1981a) Identification of the *Uvra*-gene product. *J. Mol. Biol.*, **148**: 45-62.

Sancar, G. B. (2000) Enzymatic photoreactivation: 50 years and counting. *Mutat. Res.*, **451**: 25-37.

Sancar, G. B. and Sancar, A. (1987) Structure and function of DNA photolyases. *Trends. Biochem. Sci.*, **12**: 259-261.

Sancar, G. B., Smith, F. W., Lorence, M. C., Rupert, C. S. and Sancar, A. (1984b) Sequences of the *Escherichia coli* photolyase gene and protein. *J. Biol. Chem.*, **259**: 6033-6038.

Sancar, G. B., Smith, F. W., Reid, R., Payne, G., Levy, M. and Sancar, A. (1987) Action mechanism of *Escherichia coli* DNA photolyase. 1. Formation of the enzyme-substrate complex. *J. Biol. Chem.*, **262**: 478-485.

Sartorius-Neef, S. and Pfeifer, F. (2004) *In vivo* studies on putative Shine-Dalgarno sequences of the halophilic archaeon *Halobacterium salinarum*. *Mol. Microbiol.*, **51**: 579-588.

Savage, K. N., Krumholz, L. R., Oren, A. and Elshahed, M. S. (2008) *Halosarcina pallida* gen. nov., sp. nov., a halophilic archaeon from a low-salt, sulfide-rich spring. *Int. J. Syst. Evol. Microbiol.*, **58**: 856-860.

Scheuch, S., Marschaus, L., Sartorius-Neef, S. and Pfeifer, F. (2008) Regulation of *gvp* genes encoding gas vesicle proteins in halophilic Archaea. *Arch. Microbiol.*, **190**: 333-339.

Schmid, A. K., Reiss, D. J., Kaur, A., Pan, M., King, N., Van, P. T., Hohmann, L., Martin, D. B. and Baliga, N. S. (2007) The anatomy of microbial cell state transitions in response to oxygen. *Genome. Res.*, **17**: 1399-1413.

Schul, W., et al. (2002) Enhanced repair of cyclobutane pyrimidine dimers and improved UV resistance in photolyase transgenic mice. *Embo J.*, **21**: 4719-4729.

Selby, C. P. and Sancar, A. (2006) A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. *Proc. Natl. Acad. Sci. U. S. A.*, **103**: 17696-17700.

Setlow, R. B. and Carrier, W. L. (1964) The disappearance of thymine dimers fromDNA: an error-correcting mechanism. *Proc. Natl. Acad. Sci. U. S. A.*, **51**: 226-231.

Setlow, R. B. and Setlow, J. K. (1962) Evidence that ultraviolet-induced thymine dimers in DNA cause biological damage. *Proc. Natl. Acad. Sci. U. S. A.*, **48**: 1250-1257.

Setlow, R. B., Swenson, P. A. and Carrier, W. L. (1963) Thymine dimers and inhibition of DNA synthesis by ultraviolet irradiation of cells. *Science.*, **142**: 1464-1466.

Sharma, N., Hepburn, D. and Fitt, P. S. (1984) Photoreactivation in pigmented and non-pigmented extreme halophiles. *Biochim. Biophys. Acta.*, **799**: 135-142.

Shukla, H. D. (2006) Proteomic analysis of acidic chaperones, and stress proteins in extreme halophile *Halobacterium* NRC-1: a comparative proteomic approach to study heat shock response. *Proteome. Sci.*, **4**: 6.

Sicheritz-Ponten, T. and Andersson, S. G. (2001) A phylogenomic approach to microbial evolution. *Nucleic. Acids. Res.*, **29**: 545-552.

Skophammer, R. G., Servin, J. A., Herbold, C. W. and Lake, J. A. (2007) Evidence for a gram-positive, eubacterial root of the tree of life. *Mol. Biol. Evol.*, **24**: 1761-1768.

Soltis, D. E., et al. (2004) Genome-scale data, angiosperm relationships, and 'ending incongruence': a cautionary tale in phylogenetics. *Trends. Plant. Sci.*, **9**: 477-483.

Soppa, J. (2006) From genomes to function: haloarchaea as model organisms. *Microbiology.*, **152**: 585-590.

Soppa, J., Baumann, A., Brenneis, M., Dambeck, M., Hering, O. and Lange, C. (2008) Genomics and functional genomics with haloarchaea. *Arch. Microbiol.*, **190**: 197-215.

Soppa, J., et al. (2009) Small RNAs of the halophilic archaeon *Haloferax volcanii*. *Biochem. Soc. Trans.*, **37**: 133-136.

Sprott, G. D. (1992) Structures of archaebacterial membrane lipids. *J. Bioenerg. Biomembr.*, **24**: 555-566.

Stanewsky, R., Kaneko, M., Emery, P., Beretta, B., Wager-Smith, K., Kay, S. A., Rosbash, M. and Hall, J. C. (1998) The cry(b) mutation identifies cryptochrome as a circadian photoreceptor in *Drosophila*. *Cell.*, **95**: 681-692.

Stanier, R. Y. and Van Niel, C. B. (1962) The concept of a bacterium. *Arch. Mikrobiol.*, **42**: 17-35.

Straub, J., Brenneis, M., Jellen-Ritter, A., Heyer, R., Soppa, J. and Marchfelder, A. (2009) Small RNAs in haloarchaea: Identification, differential expression and biological function. *RNA. Biol.*, **6**: 1-12.

Summers, D. (1998) Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability. *Mol. Microbiol.*, **29**: 1137-1145.

Sutherland, B. M. (1974) Photoreactivating enzyme from human leukocytes. *Nature.*, **248**: 109-112.

Takahashi, S., Nakajima, N., Saji, H. and Kondo, N. (2002) Diurnal change of cucumber CPD photolyase gene (CsPHR) expression and its physiological role in growth under UV-B irradiation. *Plant. Cell. Physiol.*, **43**: 342-349.

Takao, M., Kobayashi, T., Oikawa, A. and Yasui, A. (1989) Tandem arrangement of photolyase and superoxide dismutase genes in *Halobacterium halobium*. *J. Bacteriol.*, **171**: 6323-6329.

Tamada, T., Kitadokoro, K., Higuchi, Y., Inaka, K., Yasui, A., de Ruiter, P. E., Eker, A. P. and Miki, K. (1997) Crystal structure of DNA photolyase from *Anacystis nidulans*. *Nat. Struct. Biol.*, **4**: 887-891.

Tamanini, F., Chaves, I., Bajek, M. I. and van der Horst, G. T. J. (2007) Structure function analysis of mammalian cryptochromes. *Cold. Spring. Harb. Symp. Quant. Biol.*, **72**: 133-139

Tanida, H., Tahara, E., Mochizuki, M., Yamane, Y. and Ryoji, M. (2005) Purification, cDNA cloning, and expression profiles of the cyclobutane pyrimidine dimer photolyase of *Xenopus laevis*. *FEBS. J.* **272**: 6098-6108.

Tapingkae, W., Tanasupawat, S., Itoh, T., Parkin, K. L., Benjakul, S., Visessanguan, W. and Valyasevi, R. (2008) *Natrinema gari* sp. nov., a halophilic archaeon isolated from fish sauce in Thailand. *Int. J. Syst. Evol. Microbiol.*, **58**: 2378-2383.

Tarasov, V. Y., Besir, H., Schwaiger, R., Klee, K., Furtwangler, K., Pfeiffer, F. and Oesterhelt, D. (2008) A small protein from the bop-brp intergenic region of *Halobacterium salinarum* contains a zinc finger motif and regulates bop and crtB1 transcription. *Mol. Microbiol.*, **67**: 772-780.

Tashiro, R., Wang, A. H. and Sugiyama, H. (2006) Photoreactivation of DNA by an archaeal nucleoprotein Sso7d. *Proc. Natl. Acad. Sci. U. S. A.*, **103**: 16655-16659.

Tebbe, A., et al. (2005) Analysis of the cytosolic proteome of *Halobacterium salinarum* and its implication for genome annotation. *Proteomics.*, **5**: 168-179.

Thresher, R. J., et al. (1998) Role of mouse cryptochrome blue-light photoreceptor in circadian photoresponses. *Science.*, **282**: 1490-1494.

Todo, T., et al. (1997) Flavin adenine dinucleotide as a chromophore of the *Xenopus* (6-4)photolyase. *Nucleic. Acids. Res.*, **25**: 764-768.

Todo, T., Ryo, H., Yamamoto, K., Toh, H., Inui, T., Ayaki, H., Nomura, T. and Ikenaga, M. (1996) Similarity among the *Drosophila* (6-4)photolyase, a human photolyase homolog, and the DNA photolyase-blue-light photoreceptor family. *Science.*, **272**: 109-112.

Todo, T., Takemori, H., Ryo, H., Ihara, M., Matsunaga, T., Nikaido, O., Sato, K. and Nomura, T. (1993) A new photoreactivating enzyme that specifically repairs ultraviolet light-induced (6-4) photoproducts. *Nature.*, **361**: 371-374.

Tran-Moseman, A., Schauer, N. and De Bernardez Clark, E. (1999) Renaturation of *Escherichia coli*-derived recombinant human macrophage colony-stimulating factor. *Protein. Expr. Purif.*, **16**: 181-189.

Tsumoto, K., Ejima, D., Kumagai, I. and Arakawa, T. (2003) Practical considerations in refolding proteins from inclusion bodies. *Protein. Expr. Purif.*, **28**: 1-8.

Twellmeyer, J., Wende, A., Wolfertz, J., Pfeiffer, F., Panhuysen, M., Zaigler, A., Soppa, J., Welzl, G. and Oesterhelt, D. (2007) Microarray analysis in the archaeon *Halobacterium salinarum* strain R1. *PLoS ONE.*, **2**: e1064.

Ueda, T., Kato, A., Kuramitsu, S., Terasawa, H. and Shimada, I. (2005) Identification and characterization of a second chromophore of DNA photolyase from *Thermus thermophilus* HB27. *J. Biol. Chem.* . **280**: 36237-36243.

Valenzuela-Encinas, C., Neria-Gonzalez, I., Alcantara-Hernandez, R. J., Enriquez-Aragon, J. A., Estrada-Alvarado, I., Hernandez-Rodriguez, C., Dendooven, L. and Marsch, R. (2008) Phylogenetic analysis of the archaeal community in an alkaline-saline soil of the former lake Texcoco (Mexico). *Extremophiles.*, **12**: 247-254.

van de Vossenberg, J. L., Driessen, A. J., Grant, W. D. and Konings, W. N. (1999) Lipid membranes from halophilic and alkali-halophilic Archaea have a low H+ and Na+ permeability at high salt concentration. *Extremophiles.*, **3**: 253-257.

van Kley, H. and Hale, S. M. (1977) Assay for protein by dye binding. *Anal. Biochem.*, **81**: 485-487.

van Noort, J., Orsini, F., Eker, A., Wyman, C., de Grooth, B. and Greve, J. (1999) DNA bending by photolyase in specific and non-specific complexes studied by atomic force microscopy. *Nucleic. Acids. Res.*, **27**: 3875-3880.

van Ooyen, J. and Soppa, J. (2007) Three 2-oxoacid dehydrogenase operons in *Haloferax volcanii*: expression, deletion mutants and evolution. *Microbiology.*, **153**: 3303-3313.

Vande Berg, B. J. and Sancar, G. B. (1998) Evidence for dinucleotide flipping by DNA photolyase. *J. Biol. Chem.*, **273**: 20276-20284.

Vanhouten, B. (1990) Nucleotide excision repair in *Escherichia coli*. *Microbiol. Rev.*, **54**: 18-51.

Vreeland, R. H., Straight, S., Krammes, J., Dougherty, K., Rosenzweig, W. D. and Kamekura, M. (2002) *Halosimplex carlsbadense* gen. nov., sp. nov., a unique halophilic archaeon, with three 16S rRNA genes, that grows only in defined medium with glycerol and acetate or pyruvate. *Extremophiles.*, **6**: 445-542.

Wakagi, T., Oshima, T., Imamura, H. and Matsuzawa, H. (1998) Cloning of the gene for inorganic pyrophosphatase from a thermoacidophilic archaeon, *Sulfolobus* sp. strain 7, and overproduction of the enzyme by coexpression of tRNA for arginine rare codon. *Biosci. Biotechnol. Biochem.*, **62**: 2408-2414.

Walsby, A. E. (1980) A Square Bacterium. *Nature.*, **283**: 69-71.

Walsby, A. E. (1994) Gas vesicles. *Microbiol. Rev.*, **58**: 94-144.

Walsby, A. E. (2005) Archaea with square cells. *Trends. Microbiol.*, **13**: 193-195.

Wang, B., Jordan, S. P. and Jorns, M. S. (1988) Identification of a pterin derivative in *Escherichia coli* DNA photolyase. *Biochemistry.*, **27**: 4222-4226.

Wang, Z. G., Wu, X. H. and Friedberg, E. C. (1993) Nucleotide-excision repair of DNA in cell-free-extracts of the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.*, **90**: 4907-4911.

Wanner, C. and Soppa, J. (1999) Genetic identification of three ABC transporters as essential elements for nitrate respiration in *Haloferax volcanii*. *Genetics.*, **152**: 1417-1428.

Waters, E., et al. (2003) The genome of *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. U. S. A.*, **100**: 12984-12988.

Waugh, D. S. (2005) Making the most of affinity tags. *Trends. Biotechnol.*, **23**: 316-320.

Wende, A., Furtwangler, K. and Oesterhelt, D. (2009) Phosphate dependent Behavior of the archaeon *Halobacterium salinarum* strain R1. *J. Bacteriol.* **191**: 3852-3860.

Whitehead, K., Kish, A., Pan, M., Kaur, A., Reiss, D. J., King, N., Hohmann, L., DiRuggiero, J. and Baliga, N. S. (2006) An integrated systems approach for understanding cellular responses to gamma radiation. *Mol. Syst. Biol.*, **2**: 47.

Whitmore, L. and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers.*, **89**: 392-400.

Woese, C. R. and Fox, G. E. (1977) Phylogenetic structure of prokaryotic domain - primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.*, **74**: 5088-5090.

Woese, C. R., Gupta, R., Hahn, C. M., Zillig, W. and Tu, J. (1984) The phylogenetic relationships of three sulfur dependent archaebacteria. *Syst. Appl. Microbiol.*, **5**: 97-105.

Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.*, **87**: 4576-4579.

Wood, R. D., Robins, P. and Lindahl, T. (1988) Complementation of the Xeroderma Pigmentosum DNA-repair defect in cell-free-extracts. *Cell.*, **53**: 97-106.

Woods, W. G. and Dyall-Smith, M. L. (1997) Construction and analysis of a recombination-deficient (radA) mutant of *Haloferax volcanii*. *Mo.l Microbiol.*, **23**: 791-797.

Worthington, E. N., Kavakli, I. H., Berrocal-Tito, G., Bondo, B. E. and Sancar, A. (2003) Purification and characterization of three members of the photolyase/cryptochrome family blue-light photoreceptors from *Vibrio cholerae*. *J. Biol. Chem.*, **278**: 39143-39154.

Xue, H., Tong, K. L., Marck, C., Grosjean, H. and Wong, J. T. F. (2003) Transfer RNA paralogs: evidence for genetic code-amino acid biosynthesis coevolution and an archaeal root of life. *Gene.*, **310**: 59-66.

Xue, Y., Fan, H., Ventosa, A., Grant, W. D., Jones, B. E., Cowan, D. A. and Ma, Y. (2005) *Halalkalicoccus tibetensis* gen. nov., sp. nov., representing a novel genus of haloalkaliphilic archaea. *Int. J. Syst. Evol. Microbiol.*, **55**: 2501-2505.

Yachai, M., Tanasupawat, S., Itoh, T., Benjakul, S., Visessanguan, W. and Valyasevi, R. (2008) *Halobacterium piscisalsi* sp. nov., from fermented fish (pla-ra) in Thailand. *Int. J. Sys.t Evol. Microbiol.*, **58**: 2136-2140.

Yamada, Y., Fujiwara, T., Sato, T., Igarashi, N. and Tanaka, N. (2002) The 2.0 A crystal structure of catalase-peroxidase from *Haloarcula marismortui*. *Nat. Struct. Biol.*, **9**: 691-695.

Yang, K., Matsika, S. and Stanley, R. J. (2007) 6MAP, a fluorescent adenine analogue, is a probe of base flipping by DNA photolyase. *J. Phys. Chem.*, **111**: 10615-10625.

Yasuhira, S. and Yasui, A. (1992) Visible light-inducible photolyase gene from the goldfish *Carassius auratus*. *J. Biol. Chem.*, **267**: 25644-25647.

Yasui, A., Eker, A. P. M., Yasuhira, S., Yajima, H., Kobayashi, T., Takao, M. and Oikawa, A. (1994) A new class of DNA photolyases present in various organisms including aplacental mammals. *Embo J.*, **13**: 6143-6151.

Yasui, A. and Laskowski, W. (1975) Determination of the number of photoreactivating enzyme molecules per haploid *Saccharomyces* cells. *Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med.*, **28**: 511-518.

Zaccai, G., Cendrin, F., Haik, Y., Borochov, N. and Eisenberg, H. (1989) Stabilization of halophilic malate dehydrogenase. *J. Mol. Biol.*, **208**: 491-500.

Zaigler, A., Schuster, S. C. and Soppa, J. (2003) Construction and usage of a onefold-coverage shotgun DNA microarray to characterize the metabolism of the archaeon *Haloferax volcanii*. *Mol. Microbiol.*, **48**: 1089-1105.

Zhao, X. D., Liu, J. Q., Hsu, D. S., Zhao, S. Y., Taylor, J. S. and Sancar, A. (1997) Reaction mechanism of (6-4) photolyase. *J. Biol. Chem.*, **272**: 32580-32590.

Zillig, W., Schnabel, R. and Stetter, K. O. (1985) Archaebacteria and the origin of the eukaryotic cytoplasm. *Curr. Top. Microbiol. Immunol.*, **114**: 1-18.

Zimmermann, P. and Pfeifer, F. (2003) Regulation of the expression of gas vesicle genes in *Haloferax mediterranei*: interaction of the two regulatory proteins GvpD and GvpE. *Mol. Microbiol.*, **49**: 783-794.

Zwetsloot, J. C. M., Vermeulen, W., Hoeijmakers, J. H. J., Yasui, A., Eker, A. P. M. and Bootsma, D. (1985) Microinjected photoreactivating enzymes from *Anacystis* and *Saccharomyces* monomerize dimers in chromatin of human-cells. *Mutat. Res.*, **146**: 71-77.

A1. *Hfx. volcanii phr2* gene, flanking sequences and primer sequences (5′ - 3′)

agcgattaagtgcagtcgaacgagccggtcggccgatttttcggcccgtcggcgtccgactgccgcta
GCGC**CCATGG**AACTGCACTGGCACCG
cattcacatagccgccgtcctgaggcgcgcac**ATG**CAACTGCACTGGCACCGCCGGGACCTGCGCGTC
GCGGACAACCGCGGCCTCGCGACGGCCGCGGAGGCGGGCCCCGTCGCCCCGCTTTTCGTCTTCGATCG
GGACGTGCTCGACCACGCCGGCGCGCCCCGCGTCAGGTACTTACTCGACGCGCTCGCCGAACTCCGCG
GCGCGTACCAAGAGCGCGGGAGCGACCTGCTCGTCGCCCGCGGCGACCCCCGGACGGTCGTGCCGGCG
GTGGCGGCGGCGTTCGACGCCGAGCGGGCCGTGTGGGGTATCGACTACTCCGGACTCGCCCGCGAGCG
CGACGCCGACGTGCGGCTCGCCCTCGACGACGCCGGCGTCGCCCGCGAGCCGGTCCACGACGCCATCT
TCCACCCGCCGGGGAGCATCACGACAAACGCGGGCGACACCTACTCCGTCTACACCTACTTCTGGAAG
AAGTGGCGCGACCGCGACAAACCCGACCCCTATCCGGAACCGGACGCCGACGACCTCGTCGACGCCGC
CACGCTCGAATCCGCGGCCGAAGACCTCACCAACGGCGACGCAGAGTTCGACATCGCCGTCGGCGGCC
TCCCGACCATCTCGGACCTCGGCTTCGAGGAGCCGTCGGCATCGGTCCAGCCGGCGGGGACCGAGGCG
GCCCGCGAGCGCCTCTCGGCGTTCTGCGCGGACGCCATCTACCGCTACGCCGACGACCGCGACTACCC
GACGCGGGACGCCACGTCGCGGCTCTCGACGGACCTCAAGTTCGGGACCATCGGCATCCGCGAGGTGT
ACGCGGCGACCGCCGCGGCCCGCGAGGGCGTCGGCGGCGAGAGAGACGAGTCGGTCGAGGAGTTCCAG
TCGCAGTTGGCGTGGCGGGAGTTCTACGCCCACGTCCTCCGAGGGCATCCGAACGTCGTCACGGAGAA
CTACAAGGAGTACGAGGAGGACATCGCGTGGCGCGACGACGACGAGGAACTCGCGGCGTGGAAGGCGG
GCGAAACCGGCTACCCCATCGTCGACGCCGGGATGCGACAGCTCCGCGAGGAGGCGTACATGCACAAC
CGCGTCCGGATGATCGTCGCGTCGTTCCTCACGAAGGACCTGCTGTGCGACTGGCGACACGGCTACGC
GCACTTCCGCGAGCACCTCGCGGACCACGACACCGCAAACGACAACGGCGGCTGGCAGTGGGCCGCCT
CGACGGGGACCGATGCCCAGCCGTACTTCCGCATCTTCAACCCGATGACGCAGGGCGAGCGCTACGAC
CCCGACGGCGAGTACATCAAGCGGTACGTCCCCGAGCTCTCGGACGTGACGGCCAACACCATCCACGA
GTGGCACGAGCTGACCGACCTCGAACGCGAGCGGCTCGCGCCCGACTACCCGGAGCCCATCGTGGACC
ACGCCGAGCGCCGCGAGCGCGCGATTTCGATGTTCGAGGCCGCCCGCGGCGACGAC**TGA**gcgacgact
GACGACAAAAAACGGGAGCCAG**TTCGAA**CC
Gaccccgcctgctgtttttgccctcggtcgccgcggcacgaacgactttggcctcgggacccgattc
cgagttcatgcacgtcgctatca

The *phr2* gene sequence is shown in upper case, with ATG start codon and TGA stop codon in bold. Sequences in lower case are flanking sequences to the gene, 100 bp upstream and downstream of *phr2*. Forward primer and the corresponding nucleotide sequences are shown in red, with the *Nco*I restriction site in bold. Reverse primer and the corresponding nucleotide sequences are shown in blue, with the *Hin*dIII restriction site in bold.

A2. *Hfx. volcanii phr1* gene, flanking sequences and primer sequences (5′ - 3′)

Gagacacggtcgcggtccgtcctattccacccgccccgattccatcagccccgaatccgccgttcgac
TCTAGAACGATGTCCGAGCCAACTTC
tcaagtcacccccatcctaagcgccgccgacg**ATG**TCCGAGCCAACTTCGCTCGCGTGGTTCCGCCGC
GACCACCGTCTGCACGACAACGCGGCGTTCGCGGCCGCCTGCGACGCCGACCGCGTGGTTCCCGTCTA
CTGCGTGGACCCGCGCGAGTACGGCGACCGACCGTTCGGCGGCCCCGACTCCTTCGACTTCGAGAAGA
CCGGCGCGCACCGCGCCCGCTTCCGCCTCGAATCGCTGGCCGACCTGCGGGCGTCGCTCCGCGACCGC
GGAAGCGACCTCGTCGTCCGCGAGGGTCGCCCCGAGTCCGTCCTCCCCGAGGTGGCCGCCGCCGTCGA
CGCCGACTTCGTGACGGTCCACACGCGCCCGACACCCGAGGAGTCGCGCGTCGAATCGGCGGTCGAAA
CGGAACTTCGAGACGGCGGCGTCGAACTCCGGCGCTTCTGGGGGCACACGCTCACCCACCTCGACGGC
CTCCCGATGGCGCTTTCCGACCTCCCGGACACGTACACGACGTTTCGGAAGGCGGTCGAATTGGCCGC
CGAAGACGCCGGGACGGGCGATGCAGCAGCCGGCGACGAGCACGCGAGCGACGGCGACCCCGGCGGCC
GCGACCCGCTTCCCGAACCGACCGTCCCGCCGCTTCCCGACGACGCGCCCGACGCGGGCGACCTCCCC
GCGGTCTCGGACCTCGTCGGAACCGCGGACGCGAACGCCGACCGCGCTCCCGACGACCGCGGCGTCCT
CCCCTTCGACGGCGGCGAGACGGCCGCGCTCGACCGCGTTGAGTCGTACATCTGGGCGGGCGACCACC
TCCGCGAGTACAAGGAGGCGCGGAACGGGCTGCTCGGAGCCGACTACTCCTCGAAGTTCTCGCCGTGG
CTGAACGAGGGCTGTCTCTCGCCGCGGTACGTCAAGGCCGAGGTGGACCGCTACGAGGACCGCCGAGT
CGAAAACGACTCGACGTACTGGCTCGTATTCGAACTCCGCTGGCGCGACTTCTTCCAGTTCCAGTTCG

CCAAACACGGGAGCGACTGCTTCCACCGCGAGGGCATCCGCGAACGGACCGATATCGACTGGCGACGC
GACGACGCGCAGTTCGAGCGCTGGGCGGCGGGGAAAACTGGTATCCCCTTCGTCGACGCGAACATGCG
AGAACTGAACGCGACGGGTTACATGTCGAACCGCGGGCGGCAGAACGCGGCGTCGTTCCTCGCCAACG
ACCTCAGACTGGACTGGCGACGCGGCGCGGCCTACTTCGAGACGCGACTCGTCGACTACGACCCCGCG
TCGAACTACGGCAACTGGGCCTACATCGCGGGCGTCGGCAACGACTCGCGGGAGCGCTCGTTCGACGT
GCGCTGGCAGGCGAACCGCTACGACGAGGACGCCGAGTACGTGAAAACGTGGCTCCCCGAACTCGACG
CGTTACCCGCCGAATACGCTCACGAACCGTGGAAACTGAGCGGGGCCGAACAGGCGTCCTACGGCGTC
GCTAAAGAGGCTAATCTACG**TTCGAA**
GAGTTGGGCGTGGACTACCCCGAACCGATGGTCGATTTCTCCGAT**TAG**atgcccacgcgtcgtcggt
ctcgccgccgacacgggccacggcgaccgcgagaaccgtcacgagcgcgaacccgatgaacgcgagtc
ggggaatcgtca

The *phr1* gene sequence is shown in upper case, with ATG start codon and TGA stop codon in bold. Sequences in lower case are flanking sequences to the gene, 100 bp upstream and downstream of *phr1*. Forward primer and the corresponding nucleotide sequences are shown in red, with the *Xba*I restriction site in bold. Reverse primer and the corresponding nucleotide sequences are shown in blue, with the *Hin*dIII restriction site in bold.

A3. Phr2 protein sequence

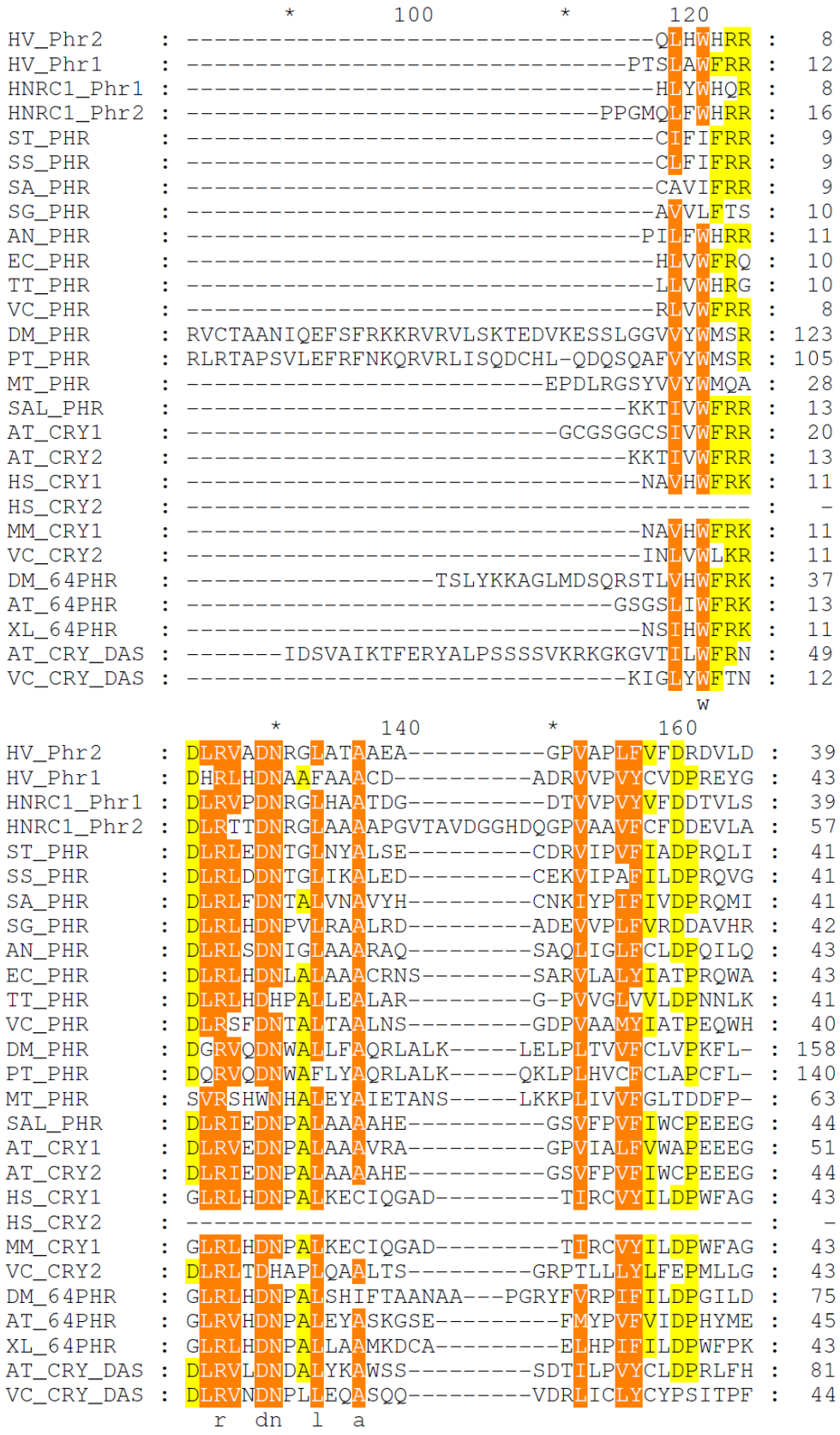
MQLHWHRRDLRVADNRGLATAAEAGPVAPLFVFDRDVLDHAGAPRVRYLLDALAELRGAYQERGSDLL
VARGDPRTVVPAVAAAFDAERAVWGIDYSGLARERDADVRLALDDAGVAREPVHDAIFHPPGSITTNA
GDTYSVYTYFWKKWRDRDKPDPYPEPDADDLVDAATLESAAEDLTNGDAEFDIAVGGLPTISDLGFEE
PSASVQPAGTEAARERLSAFCADAIYRYADDRDYPTRDATSRLSTDLKFGTIGIREVYAATAAAREGV
GGERDESVEEFQSQLAWREFYAHVLRGHPNVVTENYKEYEEDIAWRDDDEELAAWKAGETGYPIVDAG
MRQLREEAYMHNRVRMIVASFLTKDLLCDWRHGYAHFREHLADHDTANDNGGWQWAASTGTDAQPYFR
IFNPMTQGERYDPDGEYIKRYVPELSDVTANTIHEWHELTDLERERLAPDYPEPIVDHAERRERAISM
FEAARGDD

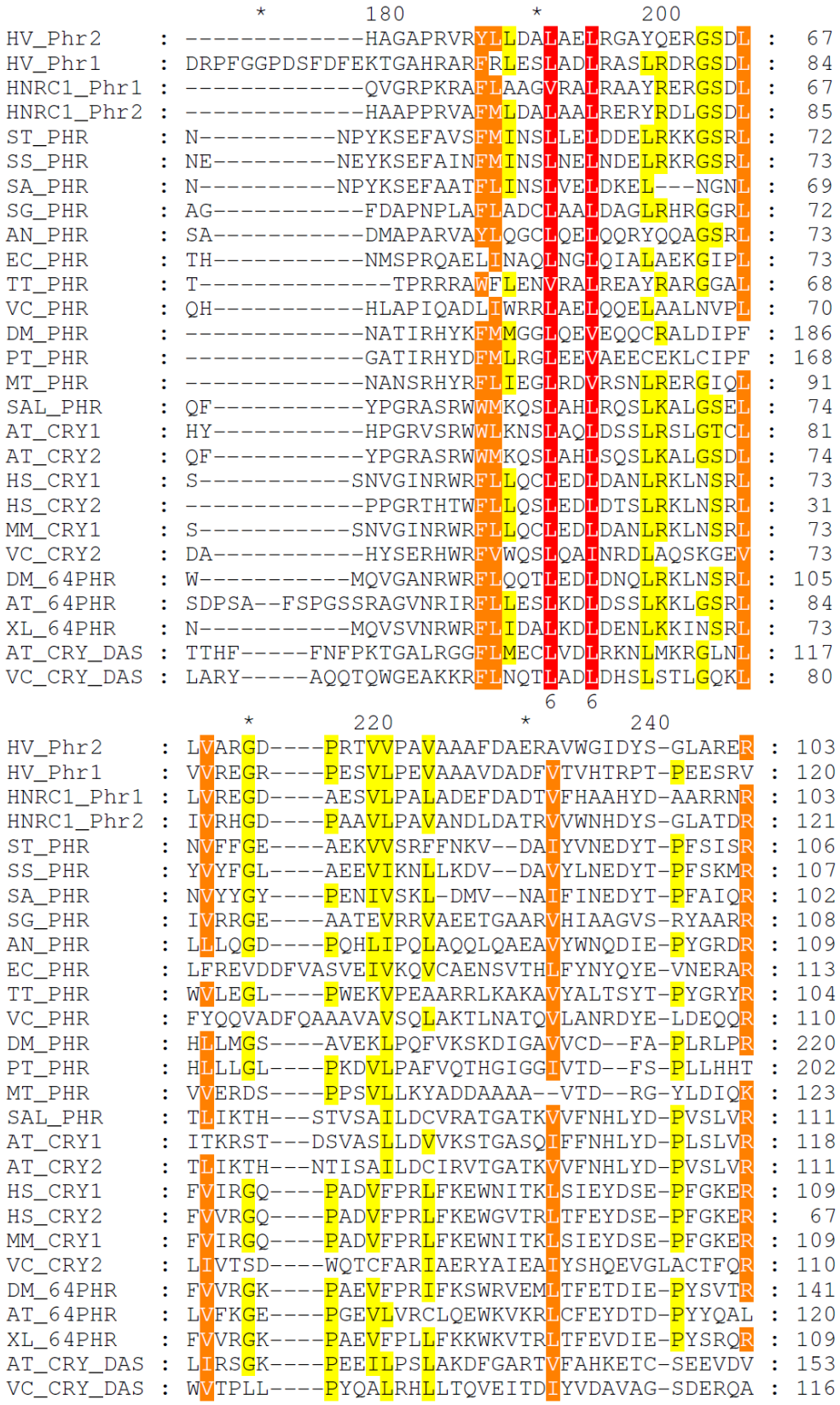
A4. Phr1 protein sequence

MSEPTSLAWFRRDHRLHDNAAFAAACDADRVVPVYCVDPREYGDRPFGGPDSFDFEKTGAHRARFRLE
SLADLRASLRDRGSDLVVREGRPESVLPEVAAAVDADFVTVHTRPTPEESRVESAVETELRDGGVELR
RFWGHTLTHLDGLPMALSDLPDTYTTFRKAVELAAEDAGTGDAAAGDEHASDGDPGGRDPLPEPTVPP
LPDDAPDAGDLPAVSDLVGTADANADRAPDDRGVLPFDGGETAALDRVESYIWAGDHLREYKEARNGL
LGADYSSKFSPWLNEGCLSPRYVKAEVDRYEDRRVENDSTYWLVFELRWRDFFQFQFAKHGSDCFHRE
GIRERTDIDWRRDDAQFERWAAGKTGIPFVDANMRELNATGYMSNRGRQNAASFLANDLRLDWRRGAA
YFETRLVDYDPASNYGNWAYIAGVGNDSRERSFDVRWQANRYDEDAEYVKTWLPELDALPAEYAHEPW
KLSGAEQASYGVELGVDYPEPMVDFSD

A5. Sequence alignment of the *Hfx. volcanii* putative Phr2 and Phr1 proteins with representative members from the different subclasses of the photolyase/crytochrome superfamily

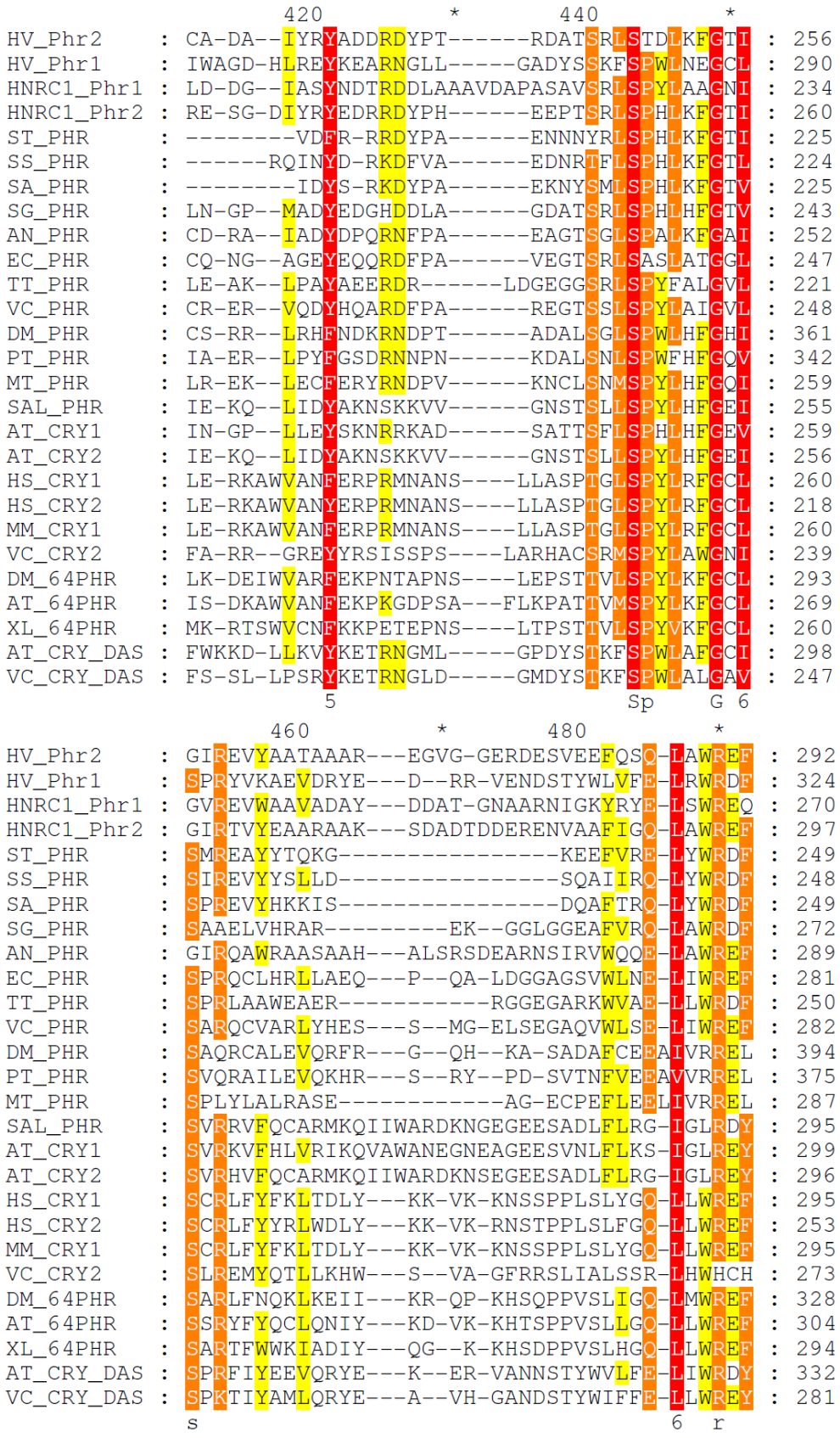
```
                          *        20         *        40
HV_Phr2     : M---------------------------------------- :    1
HV_Phr1     : MSE-------------------------------------- :    3
HNRC1_Phr1  : M---------------------------------------- :    1
HNRC1_Phr2  : MPAAQ------------------------------------ :    5
ST_PHR      : MD--------------------------------------- :    2
SS_PHR      : ML--------------------------------------- :    2
SA_PHR      : MD--------------------------------------- :    2
SG_PHR      : MSV-------------------------------------- :    3
AN_PHR      : MAA-------------------------------------- :    3
EC_PHR      : MTT-------------------------------------- :    3
TT_PHR      : MGP-------------------------------------- :    3
VC_PHR      : M---------------------------------------- :    1
DM_PHR      : MFTLASYWRESFKIVLPLQAMKRTKAQKAGPSKKAAKNEKA :   41
PT_PHR      : MDSKKRS-----------------HSTGGEAENMESQESKA :   24
MT_PHR      : MIHAERI-----------------RNLNGE----------- :   13
SAL_PHR     : MSTN------------------------------------- :    4
AT_CRY1     : MSGSVS----------------------------------- :    6
AT_CRY2     : MKMD------------------------------------- :    4
HS_CRY1     : MGV-------------------------------------- :    3
HS_CRY2     : MPA-------------------------------------- :    3
MM_CRY1     : MGV-------------------------------------- :    3
VC_CRY2     : MEK-------------------------------------- :    3
DM_64PHR    : MASWSHP-----------------QFEKGAS---------- :   14
AT_64PHR    : MAT-------------------------------------- :    3
XL_64PHR    : MRH-------------------------------------- :    3
AT_CRY_DAS  : MNDHIHRV---------------PALTEEE---------- :   15
VC_CRY_DAS  : MSK-------------------------------------- :    3
              M
                          *        60         *        80
HV_Phr2     : ---------------------------------------- :    -
HV_Phr1     : ---------------------------------------- :    -
HNRC1_Phr1  : ---------------------------------------- :    -
HNRC1_Phr2  : ---------------------------------------- :    -
ST_PHR      : ---------------------------------------- :    -
SS_PHR      : ---------------------------------------- :    -
SA_PHR      : ---------------------------------------- :    -
SG_PHR      : ---------------------------------------- :    -
AN_PHR      : ---------------------------------------- :    -
EC_PHR      : ---------------------------------------- :    -
TT_PHR      : ---------------------------------------- :    -
VC_PHR      : ---------------------------------------- :    -
DM_PHR      : SSEPKSDQESSDEEASTSKALLVSKPDYQNFEQFLTHLEHQ :   82
PT_PHR      : KRKPLQKHQFSKSNVVQKEEKDKTEGEEKGAEGLQEVVRQS :   65
MT_PHR      : ---------------------------------------- :    -
SAL_PHR     : ---------------------------------------- :    -
AT_CRY1     : ---------------------------------------- :    -
AT_CRY2     : ---------------------------------------- :    -
HS_CRY1     : ---------------------------------------- :    -
HS_CRY2     : ---------------------------------------- :    -
MM_CRY1     : ---------------------------------------- :    -
VC_CRY2     : ---------------------------------------- :    -
DM_64PHR    : ---------------------------------------- :    -
AT_64PHR    : ---------------------------------------- :    -
XL_64PHR    : ---------------------------------------- :    -
AT_CRY_DAS  : ---------------------------------------- :    -
VC_CRY_DAS  : ---------------------------------------- :    -
```

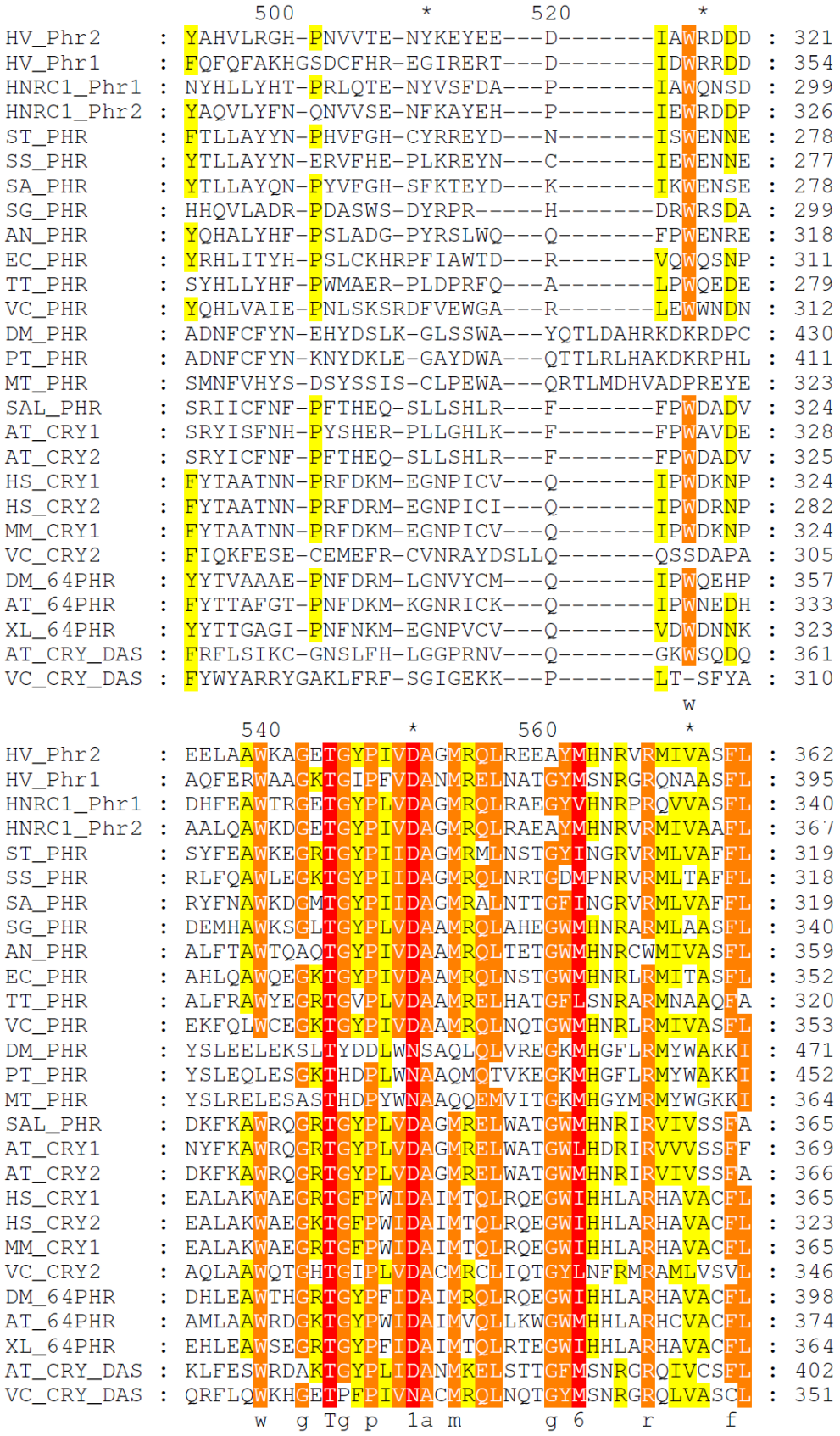

HV_Phr2 : 8
HV_Phr1 : 12
HNRC1_Phr1 : 8
HNRC1_Phr2 : 16
ST_PHR : 9
SS_PHR : 9
SA_PHR : 9
SG_PHR : 10
AN_PHR : 11
EC_PHR : 10
TT_PHR : 10
VC_PHR : 8
DM_PHR : 123
PT_PHR : 105
MT_PHR : 28
SAL_PHR : 13
AT_CRY1 : 20
AT_CRY2 : 13
HS_CRY1 : 11
HS_CRY2 : -
MM_CRY1 : 11
VC_CRY2 : 11
DM_64PHR : 37
AT_64PHR : 13
XL_64PHR : 11
AT_CRY_DAS : 49
VC_CRY_DAS : 12
HV_Phr2 : 39
HV_Phr1 : 43
HNRC1_Phr1 : 39
HNRC1_Phr2 : 57
ST_PHR : 41
SS_PHR : 41
SA_PHR : 41
SG_PHR : 42
AN_PHR : 43
EC_PHR : 43
TT_PHR : 41
VC_PHR : 40
DM_PHR : 158
PT_PHR : 140
MT_PHR : 63
SAL_PHR : 44
AT_CRY1 : 51
AT_CRY2 : 44
HS_CRY1 : 43
HS_CRY2 : -
MM_CRY1 : 43
VC_CRY2 : 43
DM_64PHR : 75
AT_64PHR : 45
XL_64PHR : 43
AT_CRY_DAS : 81
VC_CRY_DAS : 44

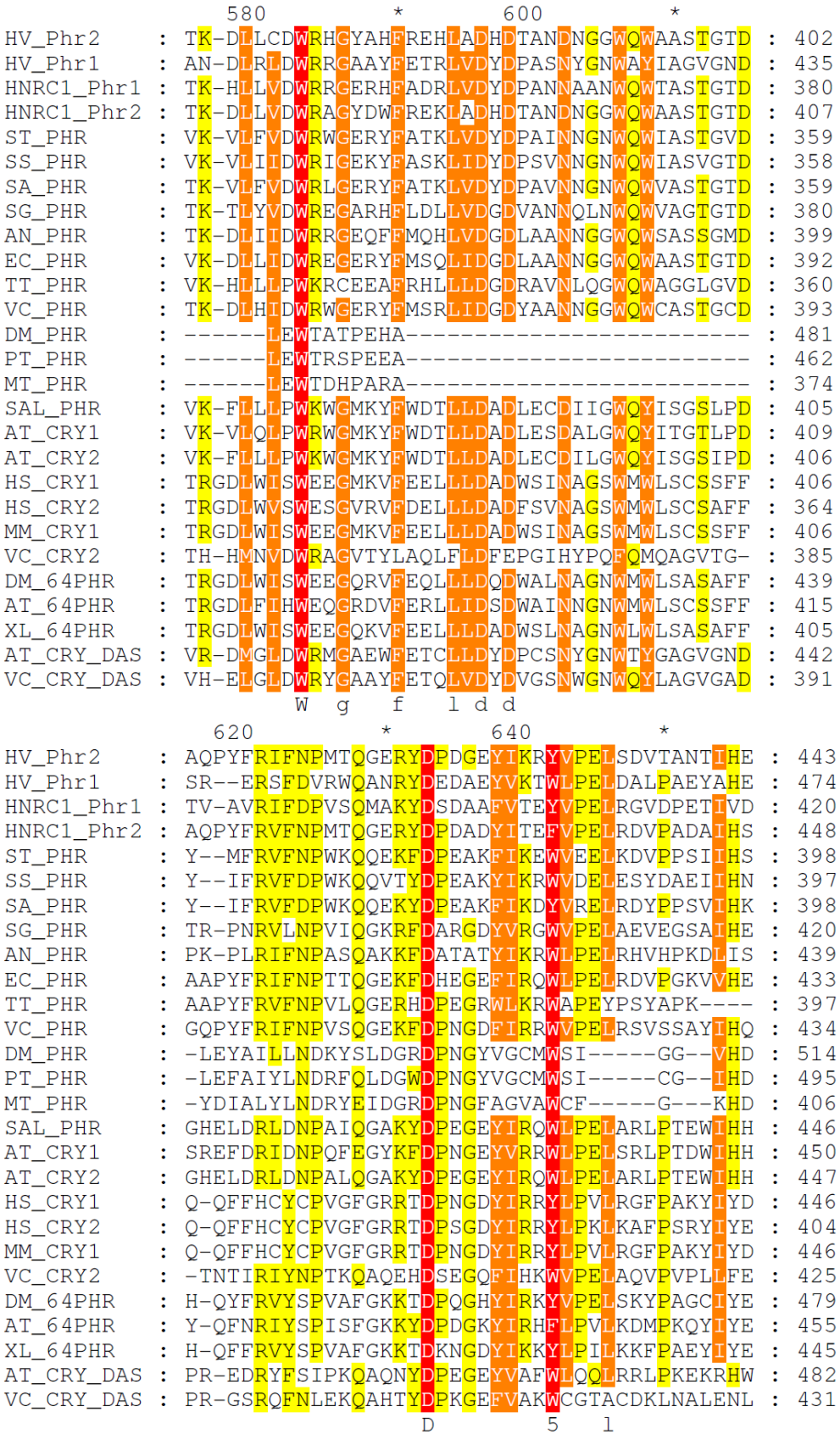

* 180 * 200
HV_Phr2 : ----------HAGAPRVRYLLDALAELRGAYQERGSDL : 67
HV_Phr1 : DRPFGGPDSFDFEKTGAHRARFRLESLADLRASLRDRGSDL : 84
HNRC1_Phr1 : ----------QVGRPKRAFLAAGVRALRAAYRERGSDL : 67
HNRC1_Phr2 : ----------HAAPPRVAFMLDALAALRERYRDLGSDL : 85
ST_PHR : N----------NPYKSEFAVSFMINSLLELDDELRKKGSRL : 72
SS_PHR : NE---------NEYKSEFAINFMINSLNELNDELRKRGSRL : 73
SA_PHR : N----------NPYKSEFAATFLINSLVELDKEL---NGNL : 69
SG_PHR : AG----------FDAPNPLAFLADCLAALDAGLRHRGGRL : 72
AN_PHR : SA----------DMAPARVAYLQGCLQELQQRYQQAGSRL : 73
EC_PHR : TH----------NMSPRQAELINAQLNGLQIALAEKGIPL : 73
TT_PHR : T----------TPRRRAWFLENVRALREAYRARGGAL : 68
VC_PHR : QH----------HLAPIQADLIWRRLAELQQELAALNVPL : 70
DM_PHR : ----------NATIRHYKFMMGGLQEVEQQCRALDIPF : 186
PT_PHR : ----------GATIRHYDFMLRGLEEVAEECEKLCIPF : 168
MT_PHR : ----------NANSRHYRFLIEGLRDVRSNLRERGIQL : 91
SAL_PHR : QF----------YPGRASRWWMKQSLAHLRQSLKALGSEL : 74
AT_CRY1 : HY----------HPGRVSRWWLKNSLAQLDSSLRSLGTCL : 81
AT_CRY2 : QF----------YPGRASRWWMKQSLAHLSQSLKALGSDL : 74
HS_CRY1 : S----------SNVGINRWRFLLQCLEDLDANLRKLNSRL : 73
HS_CRY2 : ----------PPGRTHTWFLLQSLEDLDTSLRKLNSRL : 31
MM_CRY1 : S----------SNVGINRWRFLLQCLEDLDANLRKLNSRL : 73
VC_CRY2 : DA----------HYSERHWRFVWQSLQAINRDLAQSKGEV : 73
DM_64PHR : W----------MQVGANRWRFLQQTLEDLDNQLRKLNSRL : 105
AT_64PHR : SDPSA--FSPGSSRAGVNRIRFLLESLKDLDSSLKKLGSRL : 84
XL_64PHR : N----------MQVSVNRWRFLIDALKDLDENLKKINSRL : 73
AT_CRY_DAS : TTHF-----FNFPKTGALRGGFLMECLVDLRKNLMKRGLNL : 117
VC_CRY_DAS : LARY-----AQQTQWGEAKKRFLNQTLADLDHSLSTLGQKL : 80
6 6
* 220 * 240
HV_Phr2 : LVARGD----PRTVVPAVAAAFDAERAVWGIDYS-GLARER : 103
HV_Phr1 : VVREGR----PESVLPEVAAAVDADFVTVHTRPT-PEESRV : 120
HNRC1_Phr1 : LVREGD----AESVLPALADEFDADTVFHAAHYD-AARRNR : 103
HNRC1_Phr2 : IVRHGD----PAAVLPAVANDLDATRVVWNHDYS-GLATDR : 121
ST_PHR : NVFFGE----AEKVVSRFFNKV--DAIYVNEDYT-PFSISR : 106
SS_PHR : YVYFGL----AEEVIKNLLKDV--DAVYLNEDYT-PFSKMR : 107
SA_PHR : NVYYGY----PENIVSKL-DMV--NAIFINEDYT-PFAIQR : 102
SG_PHR : IVRRGE----AATEVRRVAEETGAARVHIAAGVS-RYAARR : 108
AN_PHR : LLLQGD----PQHLIPQLAQQLQAEAVYWNQDIE-PYGRDR : 109
EC_PHR : LFREVDDFVASVEIVKQVCAENSVTHLFYNYQYE-VNERAR : 113
TT_PHR : WVLEGL----PWEKVPEAARRLKAKAVYALTSYT-PYGRYR : 104
VC_PHR : FYQQVADFQAAAVAVSQLAKTLNATQVLANRDYE-LDEQQR : 110
DM_PHR : HLLMGS----AVEKLPQFVKSKDIGAVVCD--FA-PLRLPR : 220
PT_PHR : HLLLGL----PKDVLPAFVQTHGIGGIVTD--FS-PLLHHT : 202
MT_PHR : VVERDS----PPSVLLKYADDAAAA--VTD--RG-YLDIQK : 123
SAL_PHR : TLIKTH---STVSAILDCVRATGATKVVFNHLYD-PVSLVR : 111
AT_CRY1 : ITKRST---DSVASLLDVVKSTGASQIFFNHLYD-PLSLVR : 118
AT_CRY2 : TLIKTH---NTISAILDCIRVTGATKVVFNHLYD-PVSLVR : 111
HS_CRY1 : FVIRGQ----PADVFPRLFKEWNITKLSIEYDSE-PFGKER : 109
HS_CRY2 : FVVRGQ----PADVFPRLFKEWGVTRLTFEYDSE-PFGKER : 67
MM_CRY1 : FVIRGQ----PADVFPRLFKEWNITKLSIEYDSE-PFGKER : 109
VC_CRY2 : LIVTSD----WQTCFARIAERYAIEAIYSHQEVGLACTFQR : 110
DM_64PHR : FVVRGK----PAEVFPRIFKSWRVEMLTFETDIE-PYSVTR : 141
AT_64PHR : LVFKGE----PGEVLVRCLQEWKVKRLCFEYDTD-PYYQAL : 120
XL_64PHR : FVVRGK----PAEVFPLLFKKWKVTRLTFEVDIE-PYSRQR : 109
AT_CRY_DAS : LIRSGK----PEEILPSLAKDFGARTVFAHKETC-SEEVDV : 153
VC_CRY_DAS : WVTPLL----PYQALRHLLTQVEITDIYVDAVAG-SDERQA : 116

```
                  *       260         *       280
HV_Phr2     : D--ADVRLALDDAG--VAREPVHDAI-FHPPGSIT-TNAGD : 138
HV_Phr1     : E--SAVETELRDGG--VELRRFWGHT-LTHLDGLP-MALSD : 155
HNRC1_Phr1  : R--TRVADALAAAG--VDTAGHVDHT-LVDPATLD-----D : 134
HNRC1_Phr2  : D--AGVRDALDAAG--VAHAQFHDAV-HHRPGEIR-TNAGD : 156
ST_PHR      : D--EKIRKVCEENG--IEFKAYEDYL-LTPK-SLF------ : 135
SS_PHR      : D--ERIRKYCEDTG--KIMKSFEDYL-LTSKND-F-----K : 137
SA_PHR      : D--EKMKENALNKG--IKFFSYPDVL-LSAEEKKI-----K : 133
SG_PHR      : E--QRIREALADSG--RELHVHDAVVTALAPGRVVPTGGKD : 145
AN_PHR      : D--GQVAAALKTAG--IRAVQLWDQL-LHSPDQIL-SGSGN : 144
EC_PHR      : D--VEVERALRN----VVCEGFDDSV-ILPPGAVM-TGNHE : 146
TT_PHR      : D--AKVQEALPVP-----LHLLP------APHLLP-PDLPR : 131
VC_PHR      : D--QLAQQLLSEQG--IIWSAFDDKC-VLPPGSVR-TKQGE : 145
DM_PHR      : QWVEDVGKALP-KS--VPLVQVDAHN-VVPLW-VASDKQEY : 256
PT_PHR      : QWVKDVQDALP-RQ--VPFVQVDAHN-IVPCW-VASDKQEY : 238
MT_PHR      : EWVDEAAGAL---H--IPLTQVESNV-IVPVE-TASDKEEY : 157
SAL_PHR     : D--HTVKEKLVERG--ISVQSYNGDL-CMSPGRYT-VKRAN : 146
AT_CRY1     : D--HRAKDVLTAQG--IAVRSFNADL-LYEPWEVT-DELGR : 153
AT_CRY2     : D--HTVKEKLVERG--ISVQSYNGDL-LYEPWEIY-CEKGK : 146
HS_CRY1     : D--AAIKKLATEAG--VEVIVRISHT-LYDLDKIIELNGGQ : 145
HS_CRY2     : D--AAIMKMAKEAG--VEVVTENSHT-LYDLDRIIELNGQK : 103
MM_CRY1     : D--AAIKKLATEAG--VEVIVRISHT-LYDLDKIIELNGGQ : 145
VC_CRY2     : D--LALAQWCQQHD--IVWHEFP------YAAVIR--G-AQ : 138
DM_64PHR    : D--AAVQKLAKAEG--VRVETHCSHT-IYNPELVIAKNLGK : 177
AT_64PHR    : D--VKVKDYASSTG--VEVFSPVSHT-LFNPAHIIEKNGGK : 156
XL_64PHR    : D--AEVEKLAAEHD--VQVIQKVSNT-LYDIDRIIAENNGK : 145
AT_CRY_DAS  : E--RLVNQGLKRVGNSTKLELIWGST-MYHKDDLP-FDVFD : 190
VC_CRY_DAS  : I--ARIHQDFSS----VHIHQQALHS-LLSEPQLP-FALEA : 149
                *       300         *       320
HV_Phr2     : TYSVYTYFWKKWRDR-DK----------------------PD : 157
HV_Phr1     : LPDTYTTFRKAVELAAEDAGTGDAAAGDEHASDGDPGG-RD : 195
HNRC1_Phr1  : AYPSHSQFYDDWQQA-PK----------------------PD : 153
HNRC1_Phr2  : PYSVYTYFWRKWQDR-EK----------------------NP : 175
ST_PHR      : HHRNFTSFYNEVSKV-KV----------------------RE : 154
SS_PHR      : NYRNFTTFYNVVKNK-QI----------------------KK : 156
SA_PHR      : IGRSFTPFYTKAREL-EV----------------------PK : 152
SG_PHR      : HFAVFTPYFRRWEAE-GV----------------------RG : 164
AN_PHR      : PYSVYGPFWKNWQAQ-PK----------------------PT : 163
EC_PHR      : MYKVFTPFKNAWLKR-LR------------------EGM-PE : 168
TT_PHR      : AYRVYTPFARRFLGV-E-----------------------A : 148
VC_PHR      : FFKVFTPFKRAWLTL-FQ----------------------PP : 164
DM_PHR      : AARTIRNKINSKLGE-YL--------------------SVFP : 277
PT_PHR      : GARTIRHKIHDRLPH-FL--------------------TEFP : 259
MT_PHR      : SAGTFKPKIKRHLKR-FM----------------------V : 175
SAL_PHR     : LLLVLILTGKKCLDM-SV----------------------ES : 165
AT_CRY1     : PFSMFAAFWERCLSM-PY----------------------DP : 172
AT_CRY2     : PFTSFNSYWKKCLDM-SI----------------------ES : 165
HS_CRY1     : PPLTYKRFQTLISKM-EP------------------L---EI : 165
HS_CRY2     : PPLTYKRFQAIISRM-EL------------------P---KK : 123
MM_CRY1     : PPLTYKRFQTLVSKM-EP------------------L---EM : 165
VC_CRY2     : TRKNWDEHWQQVM---R-----------------------S : 153
DM_64PHR    : APITYQKFLGIVEQL-KV------------------P---K- : 196
AT_64PHR    : PPLSYQSFLKVAGEP-SC------------------A---KS : 176
XL_64PHR    : PPLTYVRFQTVLAPL-GP------------------P---KR : 165
AT_CRY_DAS  : LPDVYTQFRKSVEAKCSI----------------------RS : 210
VC_CRY_DAS  : LPSTFTQFRKQVETI-SL----------------------SA : 168
```

```
                 *       340          *       360
HV_Phr2     : PYP-EPDADDLVDAA--TLESAAEDLTNGDAEFDIAVGGLP : 195
HV_Phr1     : PLP-EPTV---PPLPD-DAP----D----------AGDLP : 216
HNRC1_Phr1  : PVP-APDADALADV--------HDD--------------- : 169
HNRC1_Phr2  : PAP-EPEPADLAADT--ALA----D----------TSPLP : 198
ST_PHR      : PET-MEGSFDVTD----S--SMNV---------------- : 171
SS_PHR      : PVQ-NNYT-NY--YK--N--SLGD----------------- : 172
SA_PHR      : PQS-VDKSINVERIS--N--SLSI----------------- : 171
SG_PHR      : TQT-APRTVRVPDG----V---AS-------------DPLP : 184
AN_PHR      : PVA-TPT--ELVDLS--P-EQLTAI-------APLLLSELP : 191
EC_PHR      : CVA-APKVRS--------SGS---------------IEPSP : 185
TT_PHR      : PLP-APEA--LPKG--------------------------P : 160
VC_PHR      : VIG-KNRPVA-------------------------LWNVP : 178
DM_PHR      : PVVRHPHGTG--------CK----N----------VNTV- : 294
PT_PHR      : PVICHPYTSN--------V-----Q----------AEPV- : 275
MT_PHR      : PLRMRTLKMD--------SL----D----------LEPGP : 193
SAL_PHR     : -VV-LPPPWRLMPLS--A-A----E----------TVWAC : 186
AT_CRY1     : ESP-LLPPKKII------SG----D-----------VSKC : 190
AT_CRY2     : -VM-LPPPWRLMPIT--AAA----E----------AIWAC : 187
HS_CRY1     : PVE-TITSEVIEKCTTP--------LSD----DHDEKYGVP : 193
HS_CRY2     : PVG-LVTSQQMESCRAE--------IQE----NHDETYGVP : 151
MM_CRY1     : PAD-TITSDVIGKCMTP--------LSD----DHDEKYGVP : 193
VC_CRY2     : PCC-DPDL--TRAN----WLK-------------------L : 168
DM_64PHR    : VLG-V--PEKLKNMPTPPKD----EVEQ----KDSAAYDCP : 226
AT_64PHR    : --E-LVMSY--S-SLP-PIG----DI------GNLGISEVP : 200
XL_64PHR    : PIK-APTLENMKDCHTP--------WKS----SYDEKYGVP : 193
AT_CRY_DAS  : STR-IPLS--LGPTP--SVD----D----------WGDVP : 231
VC_CRY_DAS  : PMG-YPHV--L---P--PIE----Q----------GWQLP : 186
                 *       380          *       400       *
HV_Phr2     : T-ISDLGFEE----------PSA-SVQPAGTEAARERLSAF : 224
HV_Phr1     : A-VSDLVGTADANADRAPDDRG-VLPFDGGETAALDRVESY : 255
HNRC1_Phr1  : ---HTLSVPA----------PEP-PLPDAGHAAATDRLAAF : 196
HNRC1_Phr2  : S-VQELGFAE----------PEA-AVPDAGTAAARSLLDAF : 227
ST_PHR      : ----------DFLL--TFKKIES-PLFRGGRREGLYLLHRN : 199
SS_PHR      : ----------EHEL--P----PR-QGERGGRKEGIKLIERA : 196
SA_PHR      : ----------DFLL--TFKKVES-PLLKGGRPEGLMLLDRE : 199
SG_PHR      : D-RD--CVEN----------LSP-GLARGGEEAGRKLVTSW : 211
AN_PHR      : T-LKQLGFDW----------DGG-FPVEPGETAAIARLQEF : 220
EC_PHR      : S-I-TLNY-P-------RQSFDT-AHFPVEEKAAIAQLRQF : 215
TT_PHR      : E-E------GEI----PREDPGL-PLPEPGEEAALAGLRAF : 189
VC_PHR      : SALAELVWHPEQAF--DYPRIDS-TPWAADFETVRAQLRDF : 216
DM_PHR      : D-WSAAYA--SLQC--DMEVDEV-QWAKPGYKAACQQLYEF : 329
PT_PHR      : D-WNGCRA--GLQV--DRSVKEV-SWAKPGTASGLTMLQSF : 310
MT_PHR      : E-FEDAVR--DFRA--P-EDLEP-SVFRGGTSTALSIFSEF : 227
SAL_PHR     : S-VEELGLENEAEK--PSNALLT-RAWSPGWSNADKILNEF : 223
AT_CRY1     : V-ADPLVFEDDSEK--GSNALLA-RAWSPGWSNGDKALTTF : 227
AT_CRY2     : S-IEELGLENEAEK--PSNALLT-RAWSPGWSNADKLLNEF : 224
HS_CRY1     : S-LEELGFDT--------DGLSS-AVWPGGETEALTRLERH : 224
HS_CRY2     : S-LEELGFPT--------EGLGP-AVWQGGETEALARLDKH : 182
MM_CRY1     : S-LEELGFDT--------DGLSS-AVWPGGETEALTRLERH : 224
VC_CRY2     : D-AATLGLRSDIPA--TWQSKQA-GMQTGGSDMAWATLEDF : 205
DM_64PHR    : T-MKQLVKRP--------EELGP-NKFPGGETEALRRMEES : 257
AT_64PHR    : S-LEELGYKDD-------EQADW-TPFRGGESEALKRLTKS : 232
XL_64PHR    : T-LEELGQDP--------MKLGP-HLYPGGESEALSRLDLH : 224
AT_CRY_DAS  : T-LEKLGVE--------PQEVTRGMRFVGGESAGVGRVFEY : 263
VC_CRY_DAS  : L-M-DIV-----------TEPNH-SAFVGGEQAGLTHCQNY : 213
                                          g
```


420 * 440 *
HV_Phr2 : CA-DA--IYRYADDRDYPT------RDATSRLSTDLKFGTI : 256
HV_Phr1 : IWAGD-HLREYKEARNGLL-----GADYSSKFSPWLNEGCL : 290
HNRC1_Phr1 : LD-DG--IASYNDTRDDLAAAVDAPASAVSRLSPYLAAGNI : 234
HNRC1_Phr2 : RE-SG-DIYRYEDRRDYPH------EEPTSRLSPHLKFGTI : 260
ST_PHR : --------VDFR-RRDYPA------ENNNYRLSPHLKFGTI : 225
SS_PHR : ------RQINYD-RKDFVA------EDNRTFLSPHLKFGTL : 224
SA_PHR : --------IDYS-RKDYPA------EKNYSMLSPHLKFGTV : 225
SG_PHR : LN-GP--MADYEDGHDDLA------GDATSRLSPHLHFGTV : 243
AN_PHR : CD-RA--IADYDPQRNFPA------EAGTSGLSPALKFGAI : 252
EC_PHR : CQ-NG--AGEYEQQRDFPA------VEGTSRLSASLATGGL : 247
TT_PHR : LE-AK--LPAYAEERDR------LDGEGGSRLSPYFALGVL : 221
VC_PHR : CR-ER--VQDYHQARDFPA------REGTSSLSPYLAIGVL : 248
DM_PHR : CS-RR--LRHFNDKRNDPT------ADALSGLSPWLHFGHI : 361
PT_PHR : IA-ER--LPYFGSDRNNPN------KDALSNLSPWFHFGQV : 342
MT_PHR : LR-EK--LECFERYRNDPV------KNCLSNMSPYLHFGQI : 259
SAL_PHR : IE-KQ--LIDYAKNSKKVV------GNSTSLLSPYLHFGEI : 255
AT_CRY1 : IN-GP--LLEYSKNRRKAD------SATTSFLSPHLHFGEV : 259
AT_CRY2 : IE-KQ--LIDYAKNSKKVV------GNSTSLLSPYLHFGEI : 256
HS_CRY1 : LE-RKAWVANFERPRMNANS----LLASPTGLSPYLRFGCL : 260
HS_CRY2 : LE-RKAWVANYERPRMNANS----LLASPTGLSPYLRFGCL : 218
MM_CRY1 : LE-RKAWVANFERPRMNANS----LLASPTGLSPYLRFGCL : 260
VC_CRY2 : FA-RR--GREYYRSISSPS----LARHACSRMSPYLAWGNI : 239
DM_64PHR : LK-DEIWVARFEKPNTAPNS----LEPSTTVLSPYLKFGCL : 293
AT_64PHR : IS-DKAWVANFEKPKGDPSA---FLKPATTVMSPYLKFGCL : 269
XL_64PHR : MK-RTSWVCNFKKPETEPNS----LTPSTTVLSPYVKFGCL : 260
AT_CRY_DAS : FWKKD-LLKVYKETRNGML-----GPDYSTKFSPWLAFGCI : 298
VC_CRY_DAS : FS-SL-LPSRYKETRNGLD-----GMDYSTKFSPWLALGAV : 247
5 Sp G 6
460 * 480 *
HV_Phr2 : GIREVYAATAAAR---EGVG-GERDESVEEFQSQ-LAWREF : 292
HV_Phr1 : SPRYVKAEVDRYE---D--RR-VENDSTYWLVFE-LRWRDF : 324
HNRC1_Phr1 : GVREVWAAVADAY---DDAT-GNAARNIGKYRYE-LSWREQ : 270
HNRC1_Phr2 : GIRTVYEAARAAK---SDADTDDERENVAAFIGQ-LAWREF : 297
ST_PHR : SMREAYYTQKG----------KEEFVRE-LYWRDF : 249
SS_PHR : SIREVYYSLLD----------SQAIIRQ-LYWRDF : 248
SA_PHR : SPREVYHKKIS----------DQAFTRQ-LYWRDF : 249
SG_PHR : SAAELVHRAR---------EK--GGLGGEAFVRQ-LAWRDF : 272
AN_PHR : GIRQAWRAASAAH---ALSRSDEARNSIRVWQQE-LAWREF : 289
EC_PHR : SPRQCLHRLLAEQ---P--QA-LDGGAGSVWLNE-LIWREF : 281
TT_PHR : SPRLAAWEAER----------RGGEGARKWVAE-LLWRDF : 250
VC_PHR : SARQCVARLYHES---S--MG-ELSEGAQVWLSE-LIWREF : 282
DM_PHR : SAQRCALEVQRFR---G--QH--KA-SADAFCEEAIVRREL : 394
PT_PHR : SVQRAILEVQKHR---S--RY--PD-SVTNFVEEAVVRREL : 375
MT_PHR : SPLYLALRASE----------AG-ECPEFLEELIVRREL : 287
SAL_PHR : SVRRVFQCARMKQIIWARDKNGEGEESADLFLRG-IGLRDY : 295
AT_CRY1 : SVRKVFHLVRIKQVAWANEGNEAGEESVNLFLKS-IGLREY : 299
AT_CRY2 : SVRHVFQCARMKQIIWARDKNSEGEESADLFLRG-IGLREY : 296
HS_CRY1 : SCRLFYFKLTDLY---KK-VK-KNSSPPLSLYGQ-LLWREF : 295
HS_CRY2 : SCRLFYYRLWDLY---KK-VK-RNSTPPLSLFGQ-LLWREF : 253
MM_CRY1 : SCRLFYFKLTDLY---KK-VK-KNSSPPLSLYGQ-LLWREF : 295
VC_CRY2 : SLREMYQTLLKHW---S--VA-GFRRSLIALSSR-LHWHCH : 273
DM_64PHR : SARLFNQKLKEII---KR-QP-KHSQPPVSLIGQ-LMWREF : 328
AT_64PHR : SSRYFYQCLQNIY---KD-VK-KHTSPPVSLLGQ-LLWREF : 304
XL_64PHR : SARTFWWKIADIY---QG--K-KHSDPPVSLHGQ-LLWREF : 294
AT_CRY_DAS : SPRFIYEEVQRYE---K--ER-VANNSTYWVLFE-LIWRDY : 332
VC_CRY_DAS : SPKTIYAMLQRYE---A--VH-GANDSTYWIFFE-LLWREY : 281
s 6 r


500 * 520 *
HV_Phr2 : YAHVLRGH-PNVVTE-NYKEYEE---D-------IAWRDDD : 321
HV_Phr1 : FQFQFAKHGSDCFHR-EGIRERT---D-------IDWRRDD : 354
HNRC1_Phr1 : NYHLLYHT-PRLQTE-NYVSFDA---P-------IAWQNSD : 299
HNRC1_Phr2 : YAQVLYFN-QNVVSE-NFKAYEH---P-------IEWRDDP : 326
ST_PHR : FTLLAYYN-PHVFGH-CYRREYD---N-------ISWENNE : 278
SS_PHR : YTLLAYYN-ERVFHE-PLKREYN---C-------IEWENNE : 277
SA_PHR : YTLLAYQN-PYVFGH-SFKTEYD---K-------IKWENSE : 278
SG_PHR : HHQVLADR-PDASWS-DYRPR-----H-------DRWRSDA : 299
AN_PHR : YQHALYHF-PSLADG-PYRSLWQ---Q-------FPWENRE : 318
EC_PHR : YRHLITYH-PSLCKHRPFIAWTD---R-------VQWQSNP : 311
TT_PHR : SYHLLYHF-PWMAER-PLDPRFQ---A-------LPWQEDE : 279
VC_PHR : YQHLVAIE-PNLSKSRDFVEWGA---R-------LEWWNDN : 312
DM_PHR : ADNFCFYN-EHYDSLK-GLSSWA---YQTLDAHRKDKRDPC : 430
PT_PHR : ADNFCFYN-KNYDKLE-GAYDWA---QTTLRLHAKDKRPHL : 411
MT_PHR : SMNFVHYS-DSYSSIS-CLPEWA---QRTLMDHVADPREYE : 323
SAL_PHR : SRIICFNF-PFTHEQ-SLLSHLR---F-------FPWDADV : 324
AT_CRY1 : SRYISFNH-PYSHER-PLLGHLK---F-------FPWAVDE : 328
AT_CRY2 : SRYICFNF-PFTHEQ-SLLSHLR---F-------FPWDADV : 325
HS_CRY1 : FYTAATNN-PRFDKM-EGNPICV---Q-------IPWDKNP : 324
HS_CRY2 : FYTAATNN-PRFDRM-EGNPICI---Q-------IPWDRNP : 282
MM_CRY1 : FYTAATNN-PRFDKM-EGNPICV---Q-------IPWDKNP : 324
VC_CRY2 : FIQKFESE-CEMEFR-CVNRAYDSLLQ-------QSSDAPA : 305
DM_64PHR : YYTVAAAE-PNFDRM-LGNVYCM---Q-------IPWQEHP : 357
AT_64PHR : FYTTAFGT-PNFDKM-KGNRICK---Q-------IPWNEDH : 333
XL_64PHR : YYTTGAGI-PNFNKM-EGNPVCV---Q-------VDWDNNK : 323
AT_CRY_DAS : FRFLSIKC-GNSLFH-LGGPRNV---Q-------GKWSQDQ : 361
VC_CRY_DAS : FYWYARRYGAKLFRF-SGIGEKK---P-------LT-SFYA : 310
w
540 * 560 *
HV_Phr2 : EELAAWKAGETGYPIVDAGMRQLREEAYMHNRVRMIVASFL : 362
HV_Phr1 : AQFERWAAGKTGIPFVDANMRELNATGYMSNRGRQNAASFL : 395
HNRC1_Phr1 : DHFEAWTRGETGYPLVDAGMRQLRAEGYVHNRPRQVVASFL : 340
HNRC1_Phr2 : AALQAWKDGETGYPIVDAGMRQLRAEAYMHNRVRMIVAAFL : 367
ST_PHR : SYFEAWKEGRTGYPIIDAGMRMLNSTGYINGRVRMLVAFFL : 319
SS_PHR : RLFQAWLEGKTGYPIIDAGMRQLNRTGDMPNRVRMLTAFFL : 318
SA_PHR : RYFNAWKDGMTGYPIIDAGMRALNTTGFINGRVRMLVAFFL : 319
SG_PHR : DEMHAWKSGLTGYPLVDAAMRQLAHEGWMHNRARMLAASFL : 340
AN_PHR : ALFTAWTQAQTGYPIVDAAMRQLTETGWMHNRCWMIVASFL : 359
EC_PHR : AHLQAWQEGKTGYPIVDAAMRQLNSTGWMHNRLRMITASFL : 352
TT_PHR : ALFRAWYEGRTGVPLVDAAMRELHATGFLSNRARMNAAQFA : 320
VC_PHR : EKFQLWCEGKTGYPIVDAAMRQLNQTGWMHNRLRMIVASFL : 353
DM_PHR : YSLEELEKSLTYDDLWNSAQLQLVREGKMHGFLRMYWAKKI : 471
PT_PHR : YSLEQLESGKTHDPLWNAAQMQTVKEGKMHGFLRMYWAKKI : 452
MT_PHR : YSLRELESASTHDPYWNAAQQEMVITGKMHGYMRMYWGKKI : 364
SAL_PHR : DKFKAWRQGRTGYPLVDAGMRELWATGWMHNRIRVIVSSFA : 365
AT_CRY1 : NYFKAWRQGRTGYPLVDAGMRELWATGWLHDRIRVVVSSFF : 369
AT_CRY2 : DKFKAWRQGRTGYPLVDAGMRELWATGWMHNRIRVIVSSFA : 366
HS_CRY1 : EALAKWAEGRTGFPWIDAIMTQLRQEGWIHHLARHAVACFL : 365
HS_CRY2 : EALAKWAEGKTGFPWIDAIMTQLRQEGWIHHLARHAVACFL : 323
MM_CRY1 : EALAKWAEGRTGFPWIDAIMTQLRQEGWIHHLARHAVACFL : 365
VC_CRY2 : AQLAAWQTGHTGIPLVDACMRCLIQTGYLNFRMRAMLVSVL : 346
DM_64PHR : DHLEAWTHGRTGYPFIDAIMRQLRQEGWIHHLARHAVACFL : 398
AT_64PHR : AMLAAWRDGKTGYPWIDAIMVQLLKWGWMHHLARHCVACFL : 374
XL_64PHR : EHLEAWSEGRTGYPFIDAIMTQLRTEGWIHHLARHAVACFL : 364
AT_CRY_DAS : KLFESWRDAKTGYPLIDANMKELSTTGFMSNRGRQIVCSFL : 402
VC_CRY_DAS : QRFLQWKHGETPFPIVNACMRQLNQTGYMSNRGRQLVASCL : 351
w g Tg p 1a m g 6 r f


580 * 600 *
HV_Phr2 : TK-DLLCDWRHGYAHFREHLADHDTANDNGGWQWAASTGTD : 402
HV_Phr1 : AN-DLRLDWRRGAAYFETRLVDYDPASNYGNWAYIAGVGND : 435
HNRC1_Phr1 : TK-HLLVDWRRGERHFADRLVDYDPANNAANWQWTASTGTD : 380
HNRC1_Phr2 : TK-DLLVDWRAGYDWFREKLADHDTANDNGGWQWAASTGTD : 407
ST_PHR : VK-VLFVDWRWGERYFATKLVDYDPAINNGNWQWIASTGVD : 359
SS_PHR : VK-VLIIDWRIGEKYFASKLIDYDPSVNNGNWQWIASVGTD : 358
SA_PHR : TK-VLFVDWRLGERYFATKLVDYDPAVNNGNWQWVASTGTD : 359
SG_PHR : TK-TLYVDWREGARHFLDLLVDGDVANNQLNWQWVAGTGTD : 380
AN_PHR : TK-DLIIDWRRGEQFFMQHLVDGDLAANNGGWQWSASSGMD : 399
EC_PHR : VK-DLLIDWREGERYFMSQLIDGDLAANNGGWQWAASTGTD : 392
TT_PHR : VK-HLLLPWKRCEEAFRHLLLDGDRAVNLQGWQWAGGLGVD : 360
VC_PHR : TK-DLHIDWRWGERYFMSRLIDGDYAANNGGWQWCASTGCD : 393
DM_PHR : ------LEWTATPEHA------------------------ : 481
PT_PHR : ------LEWTRSPEEA------------------------ : 462
MT_PHR : ------LEWTDHPARA------------------------ : 374
SAL_PHR : VK-FLLLPWKWGMKYFWDTLLDADLECDIIGWQYISGSLPD : 405
AT_CRY1 : VK-VLQLPWRWGMKYFWDTLLDADLESDALGWQYITGTLPD : 409
AT_CRY2 : VK-FLLLPWKWGMKYFWDTLLDADLECDILGWQYISGSIPD : 406
HS_CRY1 : TRGDLWISWEEGMKVFEELLLDADWSINAGSWMWLSCSSFF : 406
HS_CRY2 : TRGDLWVSWESGVRVFDELLLDADFSVNAGSWMWLSCSAFF : 364
MM_CRY1 : TRGDLWISWEEGMKVFEELLLDADWSINAGSWMWLSCSSFF : 406
VC_CRY2 : TH-HMNVDWRAGVTYLAQLFLDFEPGIHYPQFQMQAGVTG- : 385
DM_64PHR : TRGDLWISWEEGQRVFEQLLLDQDWALNAGNWMWLSASAFF : 439
AT_64PHR : TRGDLFIHWEQGRDVFERLLIDSDWAINNGNWMWLSCSSFF : 415
XL_64PHR : TRGDLWISWEEGQKVFEELLLDADWSLNAGNWLWLSASAFF : 405
AT_CRY_DAS : VR-DMGLDWRMGAEWFETCLLDYDPCSNYGNWTYGAGVGND : 442
VC_CRY_DAS : VH-ELGLDWRYGAAYFETQLVDYDVGSNWGNWQYLAGVGAD : 391
W g f l d d
620 * 640 *
HV_Phr2 : AQPYFRIFNPMTQGERYDPDGEYIKRYVPELSDVTANTIHE : 443
HV_Phr1 : SR--ERSFDVRWQANRYDEDAEYVKTWLPELDALPAEYAHE : 474
HNRC1_Phr1 : TV-AVRIFDPVSQMAKYDSDAAFVTEYVPELRGVDPETIVD : 420
HNRC1_Phr2 : AQPYFRVFNPMTQGERYDPDADYITEFVPELRDVPADAIHS : 448
ST_PHR : Y--MFRVFNPWKQQEKFDPEAKFIKEWVEELKDVPPSIIHS : 398
SS_PHR : Y--IFRVFDPWKQQVTYDPEAKYIKRWVDELESYDAEITHN : 397
SA_PHR : Y--IFRVFDPWKQQEKYDPEAKFIKDYVRELRDYPPSVIHK : 398
SG_PHR : TR-PNRVLNPVIQGKRFDARGDYVRGWVPELAEVEGSAIHE : 420
AN_PHR : PK-PLRIFNPASQAKKFDATATYIKRWLPELRHVHPKDLIS : 439
EC_PHR : AAPYFRIFNPTTQGEKFDHEGEFIRQWLPELRDVPGKVVHE : 433
TT_PHR : AAPYFRVFNPVLQGERHDPEGRWLKRWAPEYPSYAPK---- : 397
VC_PHR : GQPYFRIFNPVSQGEKFDPNGDFIRRWVPELRSVSSAYIHQ : 434
DM_PHR : -LEYAILLNDKYSLDGRDPNGYVGCMWSI-----GG--VHD : 514
PT_PHR : -LEFAIYLNDRFQLDGWDPNGYVGCMWSI-----CG--IHD : 495
MT_PHR : -YDIALYLNDRYEIDGRDPNGFAGVAWCF-----G---KHD : 406
SAL_PHR : GHELDRLDNPAIQGAKYDPEGEYIRQWLPELARLPTEWIHH : 446
AT_CRY1 : SREFDRIDNPQFEGYKFDPNGEYVRRWLPELSRLPTDWIHH : 450
AT_CRY2 : GHELDRLDNPALQGAKYDPEGEYIRQWLPELARLPTEWIHH : 447
HS_CRY1 : Q-QFFHCYCPVGFGRRTDPNGDYIRRYLPVLRGFPAKYIYD : 446
HS_CRY2 : Q-QFFHCYCPVGFGRRTDPSGDYIRRYLPKLKAFPSRYIYE : 404
MM_CRY1 : Q-QFFHCYCPVGFGRRTDPNGDYIRRYLPVLRGFPAKYIYD : 446
VC_CRY2 : -TNTIRIYNPTKQAQEHDSEGQFIHKWVPELAQVPVPLLFE : 425
DM_64PHR : H-QYFRVYSPVAFGKKTDPQGHYIRKYVPELSKYPAGCIYE : 479
AT_64PHR : Y-QFNRIYSPISFGKKYDPDGKYIRHFLPVLKDMPKQYIYE : 455
XL_64PHR : H-QFFRVYSPVAFGKKTDKNGDYIKKYLPILKKFPAEYIYE : 445
AT_CRY_DAS : PR-EDRYFSIPKQAQNYDPEGEYVAFWLQQLRRLPKEKRHW : 482
VC_CRY_DAS : PR-GSRQFNLEKQAHTYDPKGEFVAKWCGTACDKLNALENL : 431
D 5 l

```
                660         *       680         *
HV_Phr2     : -WHELTDLE----RERLAPDYP---EPIVDHAERRERAISM : 476
HV_Phr1     : -PWKLSGAEQASYGVELGVDYP---EPM------------- : 498
HNRC1_Phr1  : -WPTLRHSQ----RERLAPDYH---HPIVDRDHAYERAQRV : 453
HNRC1_Phr2  : -WHELSLSE----RRRHAPEYP---DPIVDHSQRREDAIAM : 481
ST_PHR      : -IYKTK-----------VPGYP---SPIVNWLERVNYVKSE : 424
SS_PHR      : -AYKYT-----------LKSYP---KPIVDWRIRVNLAKRL : 423
SA_PHR      : -VYLHK-----------ITGYP---SPIVNWRERVEMVKQV : 424
SG_PHR      : -PWKLQGLD----R--AGLDYP---DPVVDLAEARAR----- : 447
AN_PHR      : -GEITPI-G--------RRGYP---APIVNHNLRQKQFKAL : 467
EC_PHR      : -PWKWAQK------AGVTLDYP---QPIVEHKEARVQTLAA : 464
TT_PHR      : ------------------------DPVVDLEEARRRYLR- : 412
VC_PHR      : -PWTYPAV--------NSVLYP---ARLVDHKQEREVTLRL : 463
DM_PHR      : MGWK----ERAIFGKVRYMNYQ---------GCRRKFDVNA : 542
PT_PHR      : QGWA----EREIFGKIRYMNYA---------GCKRKFDVAE : 523
MT_PHR      : RAWA----EREIFGKVRYMNDR---------GLKRKFRIDE : 434
SAL_PHR     : -PWDAPLTVLKASGVELGTNYA---KPIVVIDTARELLTKA : 483
AT_CRY1     : -PWNAPESVLQAAGIELGSNYP---LPIVGLDEAKARLHEA : 487
AT_CRY2     : -PWDAPLTVLKASGVELGTNYA---KPIVDIDTARELLAKA : 484
HS_CRY1     : -PWNAPEGIQKVAKCLIGVNYP---KPMVNHAEASRLNIER : 483
HS_CRY2     : -PWNAPESIQKAAKCIIGVDYP---RPIVNHAETSRLNIER : 441
MM_CRY1     : -PWNAPEGIQKVAKCLIGVNYP---KPMVNHAEASRLNIER : 483
VC_CRY2     : -PWLMTPLEAQMYQVPLESPYL---KPVMDLEASAKQARDR : 462
DM_64PHR    : -PWKASLVDQRAYGCVLGTDYP---HRIVKHEVVHKENIKR : 516
AT_64PHR    : -PWTAPLSVQTKANCIVGKDYP---KPMVLHDSASKECKRK : 492
XL_64PHR    : -PWKSPRSLQERAGCIIGKDYP---KPIVEHNVVSKQNIQR : 482
AT_CRY_DAS  : -PGRLMYMDT-------VVPLKHGNGPMAGGSKSGGGFRGS : 515
VC_CRY_DAS  : -A--LDSVD--------MVDWPI---A-------ASAYLLI : 451
                700         *       720         *       7
HV_Phr2     : FEA------------------------------------- : 479
HV_Phr1     : ------------------------------------------ :   -
HNRC1_Phr1  : FEA------------------------------------- : 456
HNRC1_Phr2  : FER------------------------------------- : 484
ST_PHR      : YKNVKAV--------------------------------- : 431
SS_PHR      : YEV------------------------------------- : 426
SA_PHR      : Y--------------------------------------- : 425
SG_PHR      : FER------------------------------------- : 450
AN_PHR      : YNQLKAA----------------------I---------- : 475
EC_PHR      : YEA------------------------------------- : 467
TT_PHR      : ------------------------------------------ :   -
VC_PHR      : Y--------------------------------------- : 464
DM_PHR      : FV-------------------------------------- : 544
PT_PHR      : ------------------------------------------ :   -
MT_PHR      : YV-------------------------------------- : 436
SAL_PHR     : ISRTREAQ----------------------I---------- : 492
AT_CRY1     : LSQMWQLE----------------------AASRAAIE-- : 503
AT_CRY2     : ISRTREAQ----------------------IMIGAAPDEI : 502
HS_CRY1     : MKQIYQQL-----------SRYR-----GL-GLLAS----- : 502
HS_CRY2     : MKQIYQQL-----------SRYR-----GL-CLLAS----- : 460
MM_CRY1     : MKQIYQQL-----------SRYR-----GL-GLLAS----- : 502
VC_CRY2     : LWQWQKLPAVQAEAMRILARHVRQAKPRTSPR---------- : 494
DM_64PHR    : MGAAYKVN-----------REVRT----------------- : 529
AT_64PHR    : MGEAYALN-----------KKM-D----GK-V---------- : 507
XL_64PHR    : MKAAYARR-----------SGSTE----GV-D---------- : 498
AT_CRY_DAS  : H----------------------------S---------- : 517
VC_CRY_DAS  : H------------------------------------- : 452
```

```
                 40          *        760          *        78
HV_Phr2      : ---------------------------------------- :    -
HV_Phr1      : ---------------------------------------- :    -
HNRC1_Phr1   : ---------------------------------------- :    -
HNRC1_Phr2   : ---------------------------------------- :    -
ST_PHR       : ---------------------------------------- :    -
SS_PHR       : ---------------------------------------- :    -
SA_PHR       : ---------------------------------------- :    -
SG_PHR       : ---------------------------------------- :    -
AN_PHR       : ---------------------------------------- :    -
EC_PHR       : ---------------------------------------- :    -
TT_PHR       : ---------------------------------------- :    -
VC_PHR       : ---------------------------------------- :    -
DM_PHR       : ---------------------------------------- :    -
PT_PHR       : ---------------------------------------- :    -
MT_PHR       : ---------------------------------------- :    -
SAL_PHR      : ---------------------------------------- :    -
AT_CRY1      : -NGSEEGLGDSAEVEEAPIEFPRDITMEETEPTRLNPNRRY : 543
AT_CRY2      : VADSFEALGANT----------------IKEPGLCPSVSS : 526
HS_CRY1      : ---------------------------------------- :    -
HS_CRY2      : ---------------------------------------- :    -
MM_CRY1      : ---------------------------------------- :    -
VC_CRY2      : ---------------------------------------- :    -
DM_64PHR     : ---------------------------------------- :    -
AT_64PHR     : ---------------------------------------- :    -
XL_64PHR     : ---------------------------------------- :    -
AT_CRY_DAS   : ---------------------------------------- :    -
VC_CRY_DAS   : ---------------------------------------- :    -
                 0          *        800          *        820
HV_Phr2      : ---------------------------------------- :    -
HV_Phr1      : ---------------------------------------- :    -
HNRC1_Phr1   : ---------------------------------------- :    -
HNRC1_Phr2   : ---------------------------------------- :    -
ST_PHR       : ---------------------------------------- :    -
SS_PHR       : ---------------------------------------- :    -
SA_PHR       : ---------------------------------------- :    -
SG_PHR       : ---------------------------------------- :    -
AN_PHR       : ---------------------------------------- :    -
EC_PHR       : ---------------------------------------- :    -
TT_PHR       : ---------------------------------------- :    -
VC_PHR       : ---------------------------------------- :    -
DM_PHR       : ---------------------------------------- :    -
PT_PHR       : ---------------------------------------- :    -
MT_PHR       : ---------------------------------------- :    -
SAL_PHR      : ---------------------------------------- :    -
AT_CRY1      : EDQMVPSIT----SSLIRPEEDEESSLNLRNSVGDSRAEVP : 580
AT_CRY2      : NDQQVPSAVRYNGSKRVKPEEEEERDM--KKSRGFDERE-- : 563
HS_CRY1      : ----VPSN---------------------------------- : 506
HS_CRY2      : ----VPSC---------------------------------- : 464
MM_CRY1      : ----VPSN---------------------------------- : 506
VC_CRY2      : ---------------------------------------- :    -
DM_64PHR     : ---------------------------------------- :    -
AT_64PHR     : ---------------------------------------- :    -
XL_64PHR     : ---------------------------------------- :    -
AT_CRY_DAS   : ---------------------------------------- :    -
VC_CRY_DAS   : ---------------------------------------- :    -
```

```
                        *       840         *       860
HV_Phr2     : ------------------------------------------ :   -
HV_Phr1     : ------------------------------------------ :   -
HNRC1_Phr1  : ------------------------------------------ :   -
HNRC1_Phr2  : ------------------------------------------ :   -
ST_PHR      : ------------------------------------------ :   -
SS_PHR      : ------------------------------------------ :   -
SA_PHR      : ------------------------------------------ :   -
SG_PHR      : ------------------------------------------ :   -
AN_PHR      : ------------------------------------------ :   -
EC_PHR      : ------------------------------------------ :   -
TT_PHR      : ------------------------------------------ :   -
VC_PHR      : ------------------------------------------ :   -
DM_PHR      : ------------------------------------------ :   -
PT_PHR      : ------------------------------------------ :   -
MT_PHR      : ------------------------------------------ :   -
SAL_PHR     : ------------------------------------------ :   -
AT_CRY1     : RNMVNTNQAQQRRAEPASNQVTAMIPEFNIRIVAES-TEDS : 620
AT_CRY2     : --LF-----------------------------------S : 566
HS_CRY1     : ----------------------PN-GNGGFMGYS-AEN- : 520
HS_CRY2     : ----------------------VEDLSHPVAE------- : 474
MM_CRY1     : ----------------------SN-GNGGLMGYAPGEN- : 521
VC_CRY2     : ------------------------------------------ :   -
DM_64PHR    : ------------------------------------------ :   -
AT_64PHR    : ------------------------------------------ :   -
XL_64PHR    : ------------------------------------------ :   -
AT_CRY_DAS  : ------------------------------------------ :   -
VC_CRY_DAS  : ------------------------------------------ :   -
                        *       880         *       900
HV_Phr2     : ------------------------------------------ :   -
HV_Phr1     : ------------------------------------------ :   -
HNRC1_Phr1  : ------------------------------------------ :   -
HNRC1_Phr2  : ------------------------------------------ :   -
ST_PHR      : ------------------------------------------ :   -
SS_PHR      : ------------------------------------------ :   -
SA_PHR      : ------------------------------------------ :   -
SG_PHR      : ------------------------------------------ :   -
AN_PHR      : ------------------------------------------ :   -
EC_PHR      : ------------------------------------------ :   -
TT_PHR      : ------------------------------------------ :   -
VC_PHR      : ------------------------------------------ :   -
DM_PHR      : ------------------------------------------ :   -
PT_PHR      : ------------------------------------------ :   -
MT_PHR      : ------------------------------------------ :   -
SAL_PHR     : ------------------------------------------ :   -
AT_CRY1     : TAESSSSG-----------------RRERSGGIVPEWSP : 642
AT_CRY2     : TAESSSSS-----------------------SVF----- : 577
HS_CRY1     : IPGCSSSG-----------------SCSQGSGIL-HYAH : 541
HS_CRY2     : -PSSSQAG-----------------SMSS-AGP------ : 488
MM_CRY1     : VPSCSSSGNGGLMGYAPGENVPSCSGGNCSQGSGIL-HYAH : 561
VC_CRY2     : ------------------------------------------ :   -
DM_64PHR    : ------------------------------------------ :   -
AT_64PHR    : ------------------------------------------ :   -
XL_64PHR    : ----KDSG-----------------Q------------- : 503
AT_CRY_DAS  : ------------------------------------------ :   -
VC_CRY_DAS  : ------------------------------------------ :   -
```

```
                        *
HV_Phr2     : ----ARGDD : 484
HV_Phr1     : ----VDFSD : 503
HNRC1_Phr1  : ---ALDGDA : 462
HNRC1_Phr2  : ----ARGDE : 489
ST_PHR      : LLEHHHHHH : 440
SS_PHR      : --CRKSKIK : 433
SA_PHR      : ----FSTKA : 430
SG_PHR      : ----ARGLD : 455
AN_PHR      : AEPEAEPDS : 484
EC_PHR      : ----ARKGK : 472
TT_PHR      : -LARDLARG : 420
VC_PHR      : ----KTAKG : 469
DM_PHR      : YGGKVHKKK : 555
PT_PHR      : FERKISPAD : 532
MT_PHR      : -RIRGLMDE : 445
SAL_PHR     : MIGACGDEM : 501
AT_CRY1     : NWRRLSQTG : 681
AT_CRY2     : TSLGKNGCK : 612
HS_CRY1     : PKVQRQSTN : 586
HS_CRY2     : TPELPSKDA : 532
MM_CRY1     : PKVQRQSSN : 606
VC_CRY2     : PNKRQPEMD : 504
DM_64PHR    : SSFEEKSET : 543
AT_64PHR    : IRNQRPKLK : 537
XL_64PHR    : SVAELFKKK : 526
AT_CRY_DAS  : GRRSRHNGP : 526
VC_CRY_DAS  : HPQNKESSS : 461
```

A6. Table of % identities

	HV Phr2	HV Phr1	HNRC1_Phr1	HNRC1_Phr2	ST_PHR	SS_PHR	SA_PHR	SG_PHR	AN_PHR	EC_PHR	TT_PHR
HV Phr2		27	42	61	28	27	27	35	38	33	28
HV Phr1			28	25	23	25	21	26	23	24	25
HNRC1_Phr1				41	28	27	27	33	32	28	29
HNRC1_Phr2					29	29	29	34	37	33	28
ST_PHR						55	60	29	29	28	25
SS_PHR							56	22	29	28	25
SA_PHR								28	27	27	24
SG_PHR									27	31	30
AN_PHR										34	30
EC_PHR											30
TT_PHR											
VC_PHR											
DM_PHR											
PT_PHR											
MT_PHR											
SAL_PHR											
AT_CRY1											
AT_CRY2											
HS_CRY1											
HS_CRY2											
MM_CRY1											
VC_CRY2											
DM_64PHR											
AT_64PHR											
XL_64PHR											
AT_CRY-DASH											
VC_CRY_DASH											

	VC_PHR	DM_PHR	PT_PHR	MT_PHR	SAL_PHR	AT_CRY1	AT_CRY2	HS_CRY1	HS_CRY2	MM_CRY1	VC_CRY2
HV Phr2	31	12	13	14	25	19	21	20	19	20	19
HV Phr1	21	11	11	14	23	17	18	19	18	19	18
HNRC1_Phr1	27	12	10	11	22	17	19	18	17	17	18
HNRC1_Phr2	33	13	13	14	27	19	23	21	21	20	18
ST_PHR	28	10	11	13	22	17	18	19	18	19	19
SS_PHR	29	10	9	10	23	16	19	19	18	19	17
SA_PHR	29	8	9	11	22	16	18	19	18	18	17
SG_PHR	30	11	11	14	25	20	20	19	18	19	18
AN_PHR	33	12	12	12	26	17	22	21	19	20	19
EC_PHR	47	13	11	13	24	19	20	19	17	18	20
TT_PHR	29	12	12	15	25	17	20	18	16	17	18
VC_PHR		12	12	13	27	19	22	20	18	19	19
DM_PHR			50	30	13	10	11	9	7	9	8
PT_PHR				32	11	10	10	10	9	10	8
MT_PHR					13	11	11	11	9	11	10
SAL_PHR						40	72	19	18	19	17
AT_CRY1							46	19	17	19	13
AT_CRY2								20	18	19	14
HS_CRY1									64	93	14
HS_CRY2										62	13
MM_CRY1											14
VC_CRY2											
DM_64PHR											
AT_64PHR											
XL_64PHR											
AT_CRY-DASH											
VC_CRY_DASH											

	DM_64PHR	AT_64PHR	XL_64PHR	AT_CRY-DASH	VC_CRY_DASH
HV Phr2	25	23	24	20	20
HV Phr1	22	23	21	37	32
HNRC1_Phr1	21	21	21	21	21
HNRC1_Phr2	24	24	23	20	19
ST_PHR	22	22	22	20	21
SS_PHR	21	22	23	21	21
SA_PHR	21	22	23	20	20
SG_PHR	23	21	22	21	21
AN_PHR	23	23	25	22	22
EC_PHR	22	23	23	20	20
TT_PHR	20	18	19	21	19
VC_PHR	23	21	23	19	22
DM_PHR	11	9	10	11	9
PT_PHR	11	11	11	10	11
MT_PHR	13	13	12	13	13
SAL_PHR	21	23	21	16	16
AT_CRY1	16	18	15	15	12
AT_CRY2	18	21	19	17	14
HS_CRY1	41	41	50	17	15
HS_CRY2	38	38	45	14	14
MM_CRY1	40	40	48	17	15
VC_CRY2	16	15	15	15	17
DM_64PHR		42	52	20	18
AT_64PHR			45	21	19
XL_64PHR				19	16
AT_CRY-DASH					31
VC_CRY_DASH					